全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



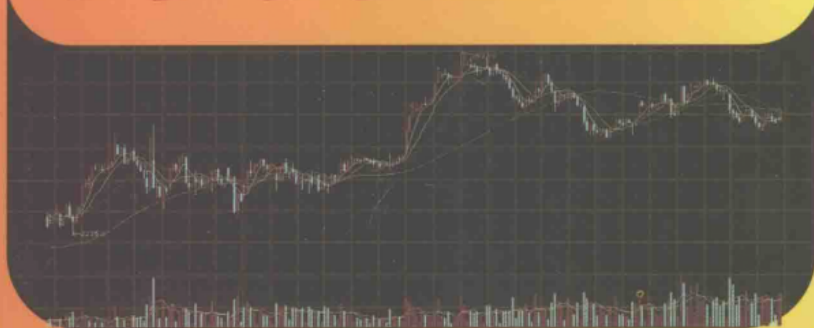
中国经图

量化投资方法丛书

刘振亚◎主编



经济计量分析基础



刘振亚
—著—

Basic Econometrics



中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net


中国经图

量化投资方法丛书

刘振亚◎主编

75% 50% 25%

经济计量分析基础

刘振亚 著


中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

图书在版编目（CIP）数据

经济计量分析基础／刘振亚著．
北京：中国经济出版社，2014．1
ISBN 978－7－5136－2830－3
Ⅰ．①经…　Ⅱ．①刘…　Ⅲ．①经济计量分析　Ⅳ．①F224．0
中国版本图书馆 CIP 数据核字（2013）第 242794 号

选题策划　张晓楹
责任编辑　李　博　李　珂
责任审读　贺　静
责任印制　张江虹
封面设计　任燕飞

出版发行　中国经济出版社
印 刷 者　三河市腾飞印务有限公司
经 销 者　各地新华书店
开　　本　800×1000mm　1/16
印　　张　15
字　　数　260 千字
版　　次　2014 年 1 月第 1 版
印　　次　2014 年 1 月第 1 次印刷
书　　号　978－7－5136－2830－3/F·9905
定　　价　36．00 元

中国经济出版社　网址 www.economyph.com　社址 北京市西城区百万庄北街 3 号　邮编 100037
本版图书如存在印装质量问题，请与本社发行中心联系调换（联系电话：010－68319116）

版权所有　盗版必究（举报电话：010－68359418　010－68319282）
国家版权局反盗版举报中心（举报电话：12390）　　服务热线：010－68344225　88386794

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

总　序

量化投资与现代科技

量化投资被广泛应用在国际对冲基金中。在过去的三十年里，由于计算机技术和统计分析技术的进步，量化投资得到了迅猛的发展。

作为量化投资基金中的杰出代表，数学家西蒙斯（Jim Simons）所领导的复兴技术（Renaissance Technology）可谓独树一帜——他的大奖章基金在1988—2008年的20年时间里创造了年均收益35.6%的奇迹。1958年，20岁的西蒙斯从麻省理工大学数学专业本科毕业后，转入加州大学伯克利分校攻读数学博士，1961年博士毕业后回母校麻省理工大学任教。一年后，他跳槽到哈佛大学任教，1964年进入美国国防分析研究院工作。1967年，西蒙斯出任纽约大学石溪分校数学系系主任。在此期间，他与著名华裔数学家陈省身合作，创造了著名的陈-西蒙斯理论。1976年，他获得美国数学学会的威布伦奖。1978年，他离开石溪大学，成为职业投资人。1988年3月，西蒙斯成立复兴技术公司。

除西蒙斯外，复兴技术的三位元老：Leonard Baum、Henry Laufer 和 James Ax 都是一流的数学家，对复兴技术的长期发展产生了很大的影响——Baum 是西蒙斯在国防分析研究院的同事，统计学中著名的 Baum - Welsh 算法的发明者，该算法被广泛应用于隐蔽的马尔科夫链、语音识别、生物和应用统计中；James Ax 也曾任石溪大学数学系主任，在数论和几何学造诣颇深，曾在1967年获得美国数学学会数论方面的科伦奖；Henry Laufer 也曾是普林斯顿大学数学教授，退休后在石溪大学创办了量化投资专业，专业课程包括：概率论与统计方法、线性规划、线性几何、数据分析、随机微积分、金融计量、最优化算法、资产定价、投资组合、金融市场、衍生品定价、固定收益产品等。

当然，成功的量化投资基金，除数学家西蒙斯所领导的复兴技术外，还有计算机教授

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

肖尔（David Shaw）领导的肖尔公司，物理学家哈丁（David Harding）领导的元胜资本（Winton Capital），经济学家格里芬（Kenneth Griffin）领导的大本营投资（Citedal）等。

一些国际知名的对冲基金对人才的要求也可归纳为以下内容：很强的电脑编程能力（C，C++；API，FIX；R，Matlab，SAS 等）；很强的数学或统计学分析技能（线性和非线性时序分析，数据挖掘，隐蔽马尔科夫链，随机分析等）；很强的大型数据库处理能力；对衍生品、资产定价、市场微结构等有深入了解。

由此可以看出，量化投资是现代金融理论、现代数理方法和现代信息技术的综合体。所涉及的金融理论主要包括：资产定价，投资组合，衍生品定价，市场微结构，行为金融学等；信息技术主要包括：编程技术（C，C++），数据库技术，交易底层通讯技术（API/FIX）；数理方法主要包括：经济计量分析基础，线性和非线性时序分析，数据挖掘，隐蔽马尔科夫链，随机分析等。

现代金融理论为量化投资提供了科学的理论基础。资产定价、衍生品定价、行为金融学和市场微结构是量化投资策略设计的科学基础；资产组合理论是量化投资降低和控制风险的有利工具。

现代信息技术和计算机技术为量化投资提供了坚实可靠的工具。软件工程和数据库技术是量化投资策略程序化的基础；API/FIX 是交易底层自动化的基础。以上两者的结合，使得量化投资能够实现完全自动化交易。有人曾说，即使西蒙斯将其所用策略公众于世，能够把公式变成钱的人在全球范围仍然是屈指可数。可见电脑技术和通讯技术在量化投资中的重要作用。

数理方法是量化投资最为核心的部分，也是数据分析和交易策略设计以及评估的基础。其所涉及的领域也是五花八门、仁者见仁、智者见智，主要包括：经济计量分析基础，线性和非线性时序分析，数据挖掘，隐蔽马尔科夫链，随机分析，小波分析等。西蒙斯曾经说过，如果抛弃了数学模型，将对我们没有任何好处。

在过去近十年的时间里，作者本人作为摩根大通期货公司（JP Morgan Futures Co.）的董事，有机会接触到一些国际知名的对冲基金，并对其进行深入的研究和合作，元胜资本（Winton Capital）就是其中的一个。作为全球最大的管理期货（期货对冲基金），元胜资本管理着约 300 亿美金的资产，在其创始人科学家哈丁先生的领导下，在 1987 年至今这 25 年的时间里创造了 16%—17% 的年均收益。仅从年均收益的数字来看，元胜资本不如复兴技术可观，但大家知道，复兴技术的大奖章基金的规模只有约 50 亿美元，而元胜资本的规模则接近 300 亿美元。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



元胜资本现有员工200多人，近50%为研究人员，远离热闹非凡的一线业务，专心从事研究工作。若简单的按照国际对冲基金2%—20%的收费标准，元胜资本一线人员创造了人均产出近亿元人民币的奇迹。可见，当今现代科技与金融的结合能够创造出巨大的财富。量化投资行业绝对是高科技产业中的佼佼者。

在过去的三十多年里，我一直从事计量经济方法和信息技术方面的学习和研究，本系列丛书也算是我和我的学生们过去多年学习和研究的小结。这里，首先要感谢我的中国老师们：中国人民大学财金学院的黄达教授（我的博士导师，前校长），陈共教授（前财政系主任），王传伦教授，经济学院的杜厚文教授（前国际经济系主任，前副校长），王景新教授（我的硕士导师），信息学院的陈禹教授（前院长），方美琪教授（我的本科导师），魏权龄教授，张怡兰教授，严颖教授，北京大学数学系的王萼芳教授。感谢我的美国老师们：George Horwich 教授（美国普渡大学，我的博士后导师），Roger Gordon 教授夫妇（加州大学圣地亚哥分校），Mark Machina 教授（加州大学圣地亚哥分校），Kajal Lahiri 教授（纽约大学阿尔巴尼分校），Harold Watts 教授（哥伦比亚大学）。他们的教育使我领略了数学的美妙、计算机的精巧、计量经济的严谨以及金融市场的奥妙，为日后的量化投资研究奠定了坚实的金融、经济、数理和计算机基础。

非常感谢中国人民大学校长陈雨露教授，金融与证券研究所所长吴晓求教授（校长助理），财金学院院长郭庆旺教授，副院长赵锡军教授，他们多年的友情、宽容和理解使得我能够有充分的自由和时间，在中国和英国两地从事我所感兴趣的研究项目。我也非常感谢我的伯明翰大学商学院同事们：David Dickinson（前院长），Nicholas Horsewood，Robert Elliott，Toby Kendall，Alessandra Guariglia 和 William Pouliot。

感谢摩根大通期货公司董事长周小雄先生多年来无微不至的兄长般关照。感谢元胜资本的创始人 David Harding，元胜亚洲 CEO Charles Allard，Kurt Settle 和田野先生。与元胜资本的合作，尤其是元胜资本每年的年会使我眼界大开，受益匪浅。感谢 AHL——牛津量化金融研究院（AHL－Oxford Institute）使我有机会参加他们所举办的世界一流水平有关量化投资的研讨会。

最后，感谢我在中国人民大学和英国伯明翰大学所指导的学生们，尤其是中国人民大学财金学院的博士生、硕士生和实验班的学生们。他们根据我的讲稿帮助整理了这套丛书中的很多内容。尤其是我的博士生杨武（现任教于中央财经大学），唐涛（现任职于中国人民银行研究所），罗涛（现任职于国家审计署），邓磊（现任职于北京工商大学），陈宇，张庆雪，刘琳，葛静，李伟，还有我在伯明翰大学的博士生曹璠玫，汪仕炫，张昆

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

（阿斯顿大学），他们对量化投资的深入研究督促着我不断学习新的知识。这些学生们对新知识的追求，使得我多年来不敢懈怠，不断学习。从他们身上，我看到了中国量化投资业的未来。

出版量化投资系列丛书的主要目的是为国内的投资者系统地介绍有关量化投资理论、技术和方法，国际上成功的量化对冲基金公司，以及量化投资的研究成果和量化投资产业的发展动向。本系列丛书内容主要包括：经济计量分析基础，投资组合理论与实践，时序分析与神经网络，金融数据挖掘，解密复兴科技，随机分析，小波分析等。

我深切地希望，此套丛书能够为中国的证券、基金、期货、私募以及个人投资者提高量化投资水平起到抛砖引玉的作用。

写这个总序前后花了我三、四个月的时间，从北京写到香港，从香港写到英国，从英国写到法国，最后又从法国写回北京。每次提笔都是千言万语涌上心头，本人最大的心愿就是在未来20到30年时间里，中国能够出现复兴技术和元胜资本这样世界一流的对冲基金。

为此，我愿奉献一生！

2013年9月于北京中国人民大学

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

前　言

在量化投资策略设计中，基本的经济计量分析方法占有举足轻重的地位。经济计量分析方法以数理统计为基础、数学方法为手段、经济理论为指导，考察经济社会中各变量之间的相互量化关系。由于经济计量分析方法具有很大的实用性，国际各大投资银行和对冲基金都不同程度地采用经济计量方法来预测金融市场的未来趋势，设计投资策略，并检验投资策略的有效性。

经济计量学的内容随时间的推进而日新月异，尤其是20世纪80年代以来，由于计算机技术的飞跃发展、个人计算机的普及和计算成本的急剧降低，人们开始更多地强调经济计量分析的应用问题，量化投资在短短的二十多年时间里，得到了突飞猛进的进步。

目前，我国有关经济计量的书籍很多是举经济学的例子，很少涉及投资领域。更为突出的问题是，由于大多数经济计量分析书籍偏重于数理推导，以数学语言代替经济语言的地方很多，这对我国大部分证券期货分析研究人员而言，学习起来十分吃力。

作者几十年来一直从事经济计量的教学和研究工作。在本书写作的过程中，作者参考了国内外大量的书刊，并与国内外专家进行了多次探讨和研究，力图在运用最少的数学语言同时，更多地强调以文字形式对基本经济计量方法加以描述证明，以便使更多的读者能够了解和应用经济计量方法。

本书在吸收了国内外大量有关经济计量著作的基础上，系统地介绍了经济计量方法及其在量化投资中的应用。在结构安排上，本书可分为以下四部分：第一部分为经济计量的数理统计基础（第一章），第二部分为经济计量分析方法（第二至十二章），第三部分为时间序列软件包 EViews（第九章），第四部分详细介绍了经济计量方法在量化投资中的应用（第十三章）。

此外，为了更好地把理论和实践相结合，本书还加以大量例子配合证明，在内容的取舍上也注重吸收国际上较为成熟的最新成果。尤其是在最后一章，在系统地介绍了经济计量分析的基础上，结合国际著名的《金融经济学杂志》2011年刊登的芝加哥大学商学院

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

T. J. Moskowitz 教授、对冲基金 AQR 资本管理公司的 Yao Hua Ooi 以及纽约大学商学院 Lasse Heje Pederson 教授的《时序动量》一文（获 2012 年度对冲基金白盒（Whitebox）研究奖），基于上证指数来阐述基本的经济计量分析方法在量化投资中的使用。

众所周知，一本数十万字的著作，它的写作过程并不是一个单纯的写作过程，期间还有大量繁重细致的劳动。从本书的第一章到最后一章，我的学生于晓筠博士（美国印地安那大学），杨武博士（中央财经大学），唐涛博士（中国人民银行研究所）等在资料收集和整理等诸多方面给了我很多的帮助，他们的友谊和帮助令我终生难忘。

希望本书能够使更多的读者应用经济计量方法去解决量化投资中的实际问题，进一步丰富量化投资的研究内容，加深研究深度，使中国的量化投资业早日腾飞。

最后应指出，由于作者水平所限，本书的缺点与错误在所难免，恳请广大读者批评指正。

于英国伯明翰大学 J. G. Smith 楼

2013 年 10 月

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

目　录

第一章　经济计量分析的统计基础 …… 1
第一节　总体与样本 …… 3
第二节　总体与样本的桥梁——随机变量 …… 6
第三节　对总体的描述——随机变量的数字特征 …… 11
第四节　对样本的描述——样本分布的数字特征 …… 14
第五节　随机变量的分布——总体和样本的连接点 …… 15
第六节　通过样本，估计总体（一）——估计量的特征 …… 22
第七节　通过样本，估计总体（二）——估计方法 …… 28
第八节　通过样本，估计总体（三）——假设检验 …… 35
第九节　应用实例：最佳持股周期 …… 42

第二章　数据与模型 …… 44
第一节　数据、模型和变量 …… 44
第二节　方程形式 …… 46
第三节　相关系数 …… 48

第三章　最小二乘法：一元线性回归 …… 51
第一节　拟合直线性质 …… 53
第二节　拟合优度的评价 …… 56
第三节　一元线性回归中的计算问题 …… 61
第四节　应用案例：对资产定价模型（CAPM）的简单检验 …… 62

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第四章　最小二乘法与多因素模型——含多个自变量的最小二乘法 ………… 66

第一节　含两个自变量的最小二乘法 ………… 66

第二节　二元回归最小二乘法解法又探 ………… 68

第三节　从另一个角度看估计系数 ………… 73

第四节　多重共线性 ………… 74

第五节　一般线性回归 ………… 75

第五章　古典线性回归模型 ………… 77

第一节　有关 ε_i 的古典模型假设 ………… 78

第二节　古典假设的一些内涵 ………… 81

第三节　高斯-马尔科夫定理 ………… 86

第四节　高斯-马尔科夫定理再探 ………… 91

第六章　正态条件下回归的推论 ………… 101

第一节　问题的引入 ………… 102

第二节　问题的预解决 ………… 102

第三节　派生的内容：自由度 ………… 103

第四节　回归系数的假设检验 ………… 105

第五节　预　测 ………… 106

第六节　复习与提高 ………… 109

第七章　假设检验 ………… 110

第一节　t 检验：对单个参数检验的方法 ………… 110

第二节　F 检验的引入 ………… 112

第三节　一般假设检验：F 检验 ………… 113

第四节　F 检验的附加例子 ………… 115

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第五节 调整后的相关系数 $\bar{R}^2$ …… 117
第六节 因果关系检验 …… 118
第七节 复习与提高 …… 122

第八章 虚拟变量 …… 124
第一节 运用虚拟变量改变回归直线的截距 …… 126
第二节 运用虚拟变量改变回归直线的斜率 …… 127
第三节 运用虚拟变量同时改变回归直线的斜率和截距 …… 128
第四节 折线回归 …… 129
第五节 复习与提高 …… 131

第九章 计量经济学软件 EViews 简介 …… 135
第一节 关于计算机的一些基本概念 …… 135
第二节 EViews 概述 …… 137
第三节 工作文件的建立及相关操作 …… 140
第四节 基本回归模型 …… 146
第五节 复习与提高 …… 147

第十章 异方差 …… 148
第一节 异方差概述 …… 150
第二节 如何发现异方差 …… 151
第三节 异方差的后果 …… 156
第四节 异方差的解决方法 …… 160

第十一章 自相关 …… 168
第一节 自相关概述 …… 168
第二节 Durbin-Watson 检验 …… 169

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第三节 自相关的处理方法 …… 170
第四节 自相关误差下的回归过程 …… 175
第五节 一阶自相关对最小二乘估计量的影响 …… 178
第六节 对含滞后因变量的序列相关模型的检验 …… 181
第七节 DW 检验的更为深入的结论 …… 182
第八节 DW 显著时的策略 …… 182

第十二章 联立方程模型 …… 185
第一节 概 述 …… 185
第二节 内生变量和外生变量 …… 188
第三节 间接最小二乘法 …… 188
第四节 识别问题 …… 190
第五节 识别的充要条件 …… 192
第六节 估计方法 …… 194
第七节 运用最小二乘法估计联立方程组问题 …… 202
第八节 复习与提高 …… 204

第十三章 基于时序动量回报率（Time – series Momentum Return）的量化交易策略 …… 206

附表 统计表 …… 211

参考书目 …… 227

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第一章 经济计量分析的统计基础

数理统计是进行经济计量分析的基本方法，即便是最为简单和实用的经济计量分析也涉及到诸多数理统计的基本概念。同时，数理统计在大学教学中属比较困难的部分。鉴于此，在进一步讨论如何进行经济计量分析之前，我们将首先对本书中经常使用到的数理统计的基本内容进行一些回顾。

事实上，不懂得数理统计就不可能学习和研究经济计量分析，数理统计是经济计量分析的基础，它为经济计量分析提供了唯一而有效的方法。这一点，读者们将在今后各章节的学习中深刻地体会到。从某种意义上来说，经济计量分析就是使数理统计在经济和投资中得以应用的一门学科。正因为如此，在世界各权威性的经济计量分析著作中，对数理统计基础内容的回顾和复习往往占据极大的篇幅，并以独立、完整而重要的章节出现。这一点从 William H. Greene 所著的《Econometric analysis》一书中可见一斑。

本书将以一个整章的篇幅对数理统计的基础原理给出一个概略的复习和回顾。应当指出的是，这一回顾是“概略的”，但不是“肤浅的”。通过本章的学习，读者将能从逻辑结构的层次上，比较完整地掌握数理统计的基本内容。无论读者事先学过还是没有学过数理统计，本章内容都会很有帮助，这不仅由于本章内容完整而自成体系，而且因为它注重于逻辑结构分析，注重于方法的阐述，注重于公式、定义和定理的内在涵义及其相互关系的揭示，同时，也注意了紧扣经济计量分析这一主题。

为了把注意力集中在数理统计的逻辑结构及其在量化投资研究的应用上，而不是数学推导的具体细节上，并且鉴于这不是一部数学书，我们对公式、定理往往强调它们的结论、涵义和它所给出的方法，而不强调定理、公式本身的数学证明。

在本章的讨论展开之前，首先给出这部分内容的逻辑结构。读者可以在以后各节的学习中随时参阅此表，以便了解自己所阅读的部分在整个数理统计体系中所处的地位，及其与其它部分的联系。在本章结束时，将再次仔细地讨论此表，并以此作为全章小结。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

数理统计学的逻辑结构

1. 总体（Population）和样本（Sample）

如何用一个变量来描述总体——随机变量（Random Variable）的引入。

2. 对总体的描述：随机变量的数字特征

数学期望 $\mu_x = E(X)$ 描述总体的一般水平

方差 $\sigma_x^2 = Var(X)$ 描述总体的离散程度

3. 对样本的描述：样本分布的数字特征

样本平均数 $\bar{X}$ 描述样本的一般水平

样本方差 S^2 描述样本的离散程度

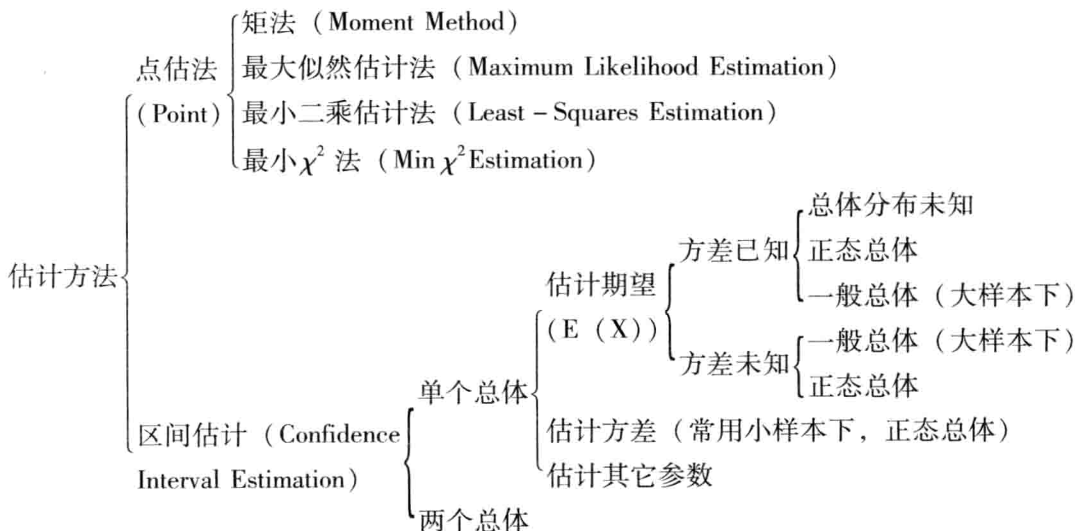
实际上，$\bar{X}$ 和 S^2 是 $\mu_x = E(X)$ 和 $\sigma_x^2 = Var(X)$ 的无偏估计量。

4. 总体与样本的连接点：随机变量的分布（Distribution）

5. 通过样本数据和样本分布特征来估计总体的数字特征，即 μ_x 和 σ_x^2，以及总体中数据生成过程的各种参数。

a. 估计量的特性

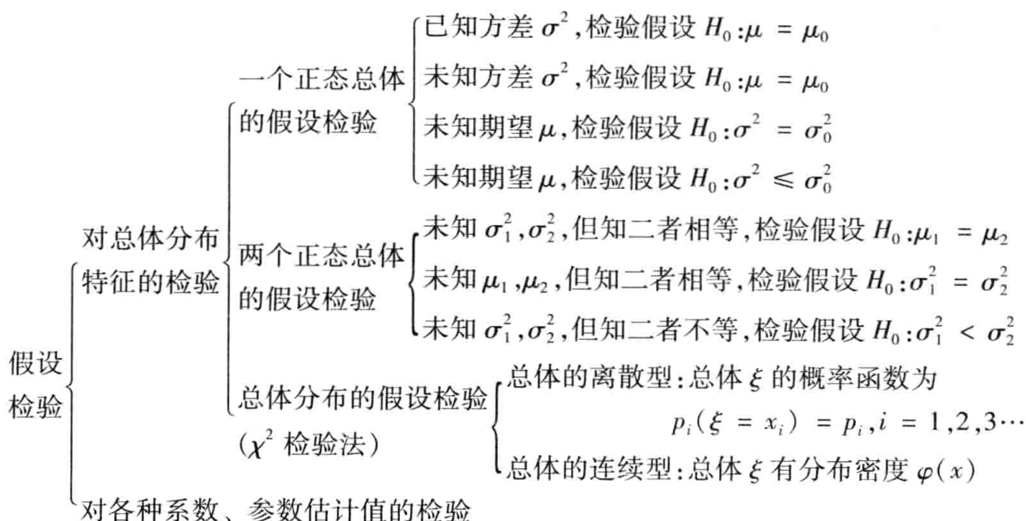
b. 估计的方法



c. 对估计量的检验——假设检验

它与区间估计本质相同，只是所讨论的出发点不同：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第一节　总体与样本

首先，我们给出三个定义：

定义 1.1.1　研究对象的全体称为总体或母体，组成总体的每个基本单位称为个体。

注意： 1. 总体根据所包含的个体数目可分为：有限总体和无限总体。

2. 总体中的每一个体，具有共同的可观察特征，也就是说，有相同的数据产生过程。我们把这种特征作为不同总体的区别指标。

3. 量度同一对象所得到的数据，也构成一种类型的总体。这种总体中的数据之所以不同，是因为测量误差的存在。

4. 表现总体状况的某一个或某几个数量特征在总体中依不同概率取不同值。

定义 1.1.2　总体中抽出若干个体而组成的集合，称为样本。样本中所含个体的个数，称为样本容量。

注意： 在进行抽样时，样本的选取必须是随机的，即总体中每一个体有同样的机会被选入样本。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

定义 1.1.3 **根据概率的不同而取不同数值的变量，叫做随机变量（Random Variable）。**

注意：1. 一个随机变量具备下列特性：它可以取到许多不同的值，每个值都可能以一个小于或等于1的概率被取到。显然，所有这些概率之和为1。

2. 随机变量以一定的概率取到各种可能值，按其取值情况可以把随机变量分为两类：离散型随机变量和连续型随机变量。离散型随机变量所可能的取值至多为可列个。连续型随机变量所可能的取值可以是整个数轴或数轴上的某个连续区间，如图1.1所示。

3. 本书中，随机变量用 x,y,ξ,η 等符号表示。

以上我们给出了总体、样本、随机变量三个重要概念，下面我们讨论三者之间的关系。

由上面我们已经了解到，表示总体状况的某一个或某几个数量特征在总体中往往随概率不同而取不同的值（定义1.1，注意4）。显然，对于这样的数量特征，我们用一般的变量是无法加以描述的，能够对之给予描述的是一类特殊的变量，即随机变量（定义1.3）。

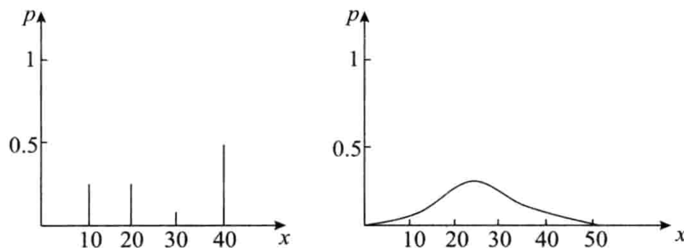


图 1.1

我们对于总体中某一个体所具有的特殊属性往往并不关心，真正感兴趣的是表示总体特征的数量指标。例如，基于股价计算出来的股票回报率等。就总体的某一数量特征 ξ 而言，如股票回报率分布，每个个体的取值不一定完全相同，但它是按照一定规律分布的，是基本稳定的。因此，对于一个总体来说，其每一个数量特征完全符合定义1.3，是根据概率的不同而取不同值的变量，即总体中每一个数量特征就是一个随机变量。由于我们主要是研究总体的数量特征，所以我们把总体看成是具有若干数量特征的研究对象的全体，可直接用一个随机变量来表示。

可见，所谓总体就是一个随机变量，所谓样本就是 n 个（样本容量为 n ）相互独立且

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

与总体有相同分布的随机变量 $x_1,\cdots,x_n$。每一次具体抽样所得的数据，就是 n 元随机变量的一个样本观测值，记为 $(X_1,\cdots,X_n)$。

任何随机变量都有自己的分布，我们已经了解总体可表示为一个随机变量，并具有其自身的分布。而样本就是 n 个相互独立且与总体有相同分布的随机变量 $x_1,\cdots,x_n$。可见，通过总体的分布可以把样本与总体联系起来，从而使我们能够通过对样本特征的研究达到对总体研究的目的，换而言之，总体分布（Distriution）是样本和总体的连接点。

每当提到一个容量为 n 的样本时，常有双重涵义：一是指某一次抽样的具体数值 $(X_1,\cdots,X_n)$；二是泛指一次抽出的可能结果，这时指一个 n 元随机变量 $(x_1,x_2,\cdots,x_n)$。

定义 1.1.4 **设 $(x_1,x_2,\cdots,x_n)$ 为一组样本，函数 $f(x_1,x_2,\cdots,x_n)$ 若不含有未知参数，则称 $f(x_1,x_2,\cdots,x_n)$ 为统计量。**

统计量一般是样本的连续函数。由于样本是随机变量，因而它的函数也是随机变量，如 $s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ 就是统计量。

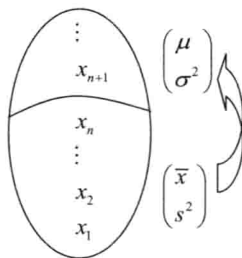


图 1.2

以上我们讨论了总体、样本和随机变量的关系，下面我们讨论总体和样本二者之间的关系。如图 1.2 所示，样本是总体的一小部分，是对总体进行随机抽取后所得到的集合。对于观察者来讲，整个总体的状况是不了解的，观察者所能了解的只是总体的一部分——样本的具体状况。我们所要做的就是通过对这些样本具体状况的研究，来推知整个总体的状况。图 1.2 中的箭头表示了我们的研究方向。既然要通过研究样本来推知总体的状况特性，那么我们通过什么能把样本和总体联系起来，使我们的研究在二者之间得以沟通呢？答案是通过对总体的数据产生过程的指定和分析，也就是说通过对随机变量函数的定义、分类和分布来进行的。随机变量是对母体和样本产生过程的数学描述，是连结母体和样本的桥梁。

下面，我们将讨论以下三个问题：随机变量的分布、二元随机变量和独立性问题。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第二节 总体与样本的桥梁——随机变量

一、随机变量的分布

（一）离散型随机变量的分布

定义 1.2.1 如果随机变量 ξ 只取有限个或可列个可能值，而且 ξ 以确定的概率取到这些值，则称 ξ 为离散型随机变量。

直观上，可将 ξ 的可能取值及相应概率列成概率分布表（见表 1－1）。

表 1－1

ξ_i	x_1	x_2	$\cdots$	x_i	$\cdots$
p	p_1	p_2	$\cdots$	p_i	$\cdots$

同时，ξ 的概率分布情况也可以用一系列等式表示：

$$p_i(\xi = x_i) = p_i \qquad (i = 1,2,\cdots)$$

此式的涵义为 ξ 取值为 x_i 的概率为 $p_i(i = 1,2,\cdots)$，它称为随机变量 ξ 的概率函数（概率分布）。这里的 $x_i(i = 1,2,\cdots)$ 应包括 ξ 的每一个可能的取值，而且每一个可能的取值都只出现一次。显然，

$$\begin{cases} p_i \geq 0 \\ \sum p_i = 1 \end{cases}$$

所谓离散型随机变量的分布，一般就是指它的概率函数或分布表。

若已知股价下跌的概率为 55%。那么，我们可以用随机变量来指述这一过程。以 $\xi = 0$ 表示股价下跌，$\xi = 1$ 表示股价上涨，其分布如表 1－2 所示①：

表 1－2

ξ_i	0	1
p	0.55	0.45

其概率函数为 $p(\xi = 0) = 0.55, p(\xi = 1) = 0.45$，或写作：$p(\xi = i) = (0.45)^i(0.55)^{1-i} \quad (i = 0,1)$。

① 注：由于股价几乎不可能出现完全不变的情况，这里对此种情况我们忽略不计。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



（二）随机变量的分布函数

定义 1.2.2 若 ξ 是一个随机变量（可以是离散的，也可以为非离散的），对任何实数 x，令 $F(x) = p(\xi \le x)$，称 $F(x)$ 为随机变量 ξ 的分布函数。

$F(x)$，即事件“$\xi \le x$”的概率，是 x 的一个实函数。对任意实数 $x_1 < x_2$，有：$p(x_1 < \xi \le x_2) = p(\xi \le x_2) - p(\xi \le x_1) = F(x_2) - F(x_1)$。可见，若已知 ξ 的分布函数 $F(x)$，就能知道 ξ 在任一个区间上取值的概率。所以，它完整地描述了随机变量的变化情况。

$F(x)$ 的性质：

1. 对一切 $x \in (-\infty, +\infty)$，$0 \le F(x) \le 1$；
2. $F(x)$ 为非减函数；
3. $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0$，$F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1$；
4. $F(x)$ 至多有可列个间断点，且在间断点上右连续。

若还以前述股价变化为例，则其分布函数应为：

$$F(x) = p(\xi \le x) = \begin{cases} 0 & x < 0 \\ 0.55 & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

（三）连续型随机变量的分布

定义 1.2.3 对于任何实数 x，如果随机变量 ξ 的分布函数 $F(x)$ 可以写成 $F(x) = \int_{-\infty}^{x} \varphi(t)dt$，其中 $\varphi(x) \ge 0$，则称其为连续型随机变量，称 $\varphi(x)$ 为 ξ 的概率分布密度函数，也常写为 $\xi \sim \varphi(x)$。

$\varphi(x)$ 的性质：

1. $\varphi(x) \geqslant 0$；2. $\int_{-\infty}^{+\infty} \varphi(x)dx = 1$。

显然 $P(a < \xi \leqslant b) = \int_{a}^{b} \varphi(x)dx$，并且在 $\varphi(x)$ 的连续点上，有 $F'(x) = \varphi(x)$。

下面我们来看 $\varphi(x)$ 的涵义：

$\because F'(x) = \varphi(x)$

$$\therefore \varphi(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{p(x < \xi < x + \Delta x)}{\Delta x}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



这表明，$\varphi(x)$ 不是 ξ 取值 x 的概率，而是在 x 点概率分布的密集程度。但是 $\varphi(x)$ 的大小能反映出 ξ 在 x 附近取值的概率大小。在本书中，我们有时直接把 $\varphi(x)$ 看成 ξ 取值 x 的概率。当然，这是不严谨的。

例 若 ξ 有密度函数 $\varphi(x)=\begin{cases}\lambda & a\leq x\leq b \quad (a<b)\\ 0 & 其他\end{cases}$，则称 ξ 服从区间 $[a,b]$ 上的均匀分布。试求 $F(x)$。

解 $\because \int_{-\infty}^{+\infty}\varphi(x)dx=\int_a^b\lambda dx=\lambda(b-a)=1$

$\therefore \lambda=\frac{1}{b-a}$

$\because F(x)=\int_{-\infty}^{x}\varphi(t)dt$

$$\therefore F(x)=\begin{cases}0 & x<a\\ \frac{x-a}{b-a} & a\leq x\leq b\\ 1 & b\leq x\end{cases}$$

（四）分布函数、概率函数、密度函数三者的关系

分布函数既适用于离散型随机变量，也可适用于连续型随机变量，是描述各种类型随机变量变化规律的最一般的共同形式。但是由于它不够直观，所以往往不常用。对于离散型随机变量，概率函数的描述既简单又直观；对于连续型随机变量，其密度函数 $\varphi(x)$ 的大小能反映出 ξ 在 x 附近取值的概率大小，从而要比分布函数的描述直观。因而在实际应用中，一般用概率函数和密度函数来分别对离散型和连续型的随机变量进行描述。

二、二元随机变量

定义 1.2.4 如果每次试验的结果对应着一组确定的实数 $(\xi_1,\cdots,\xi_n)$，它们是随试验结果不同而变化的 n 个随机变量。如果对任意一组实数 $x_1,\cdots,x_n$，事件“$\xi_1\leq x_1,\cdots,\xi_n\leq x_n$”有着确定的概率，则称 n 个随机变量的总体 $(\xi_1,\cdots,\xi_n)$ 为一个 n 元随机变量。

定义 1.2.5 称 n 元函数 $F(x_1,\cdots,x_n)=p(\xi_1\leq x_1,\cdots,\xi_n\leq x_n)$，$(x_1,\cdots,x_n)\in R^n$ 为 n 元随机变量的分布函数

定义 1.2.6 如果二元随机变量 (ξ,η) 所有可能的取值为有限或可列个，并且以确定的概率取各个不同数值，则称 (ξ,η) 为二元离散型随机变量。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

为了直观，可以把 (ξ,η) 所有可能的取值及相应概率列成下表，称为 (ξ,η) 的概率分布表，如表 1－3 所示。

表 1－3

ξ \ η	y_1	y_2	…	y_j	…
x_1	p_{11}	p_{12}	…	p_{1j}	…
x_2	p_{21}	p_{22}	…	p_{2j}	…
…	…	…	…	…	…
x_i	p_{i1}	p_{i2}	…	p_{ij}	…
…	…	…	…	…	…

为了简单，也可以用一系列等式表示离散型二元随机变量 (ξ,η) 的概率分布，即：

$$p(\xi = x_i,\eta = y_i) = p_{ij} \qquad (i,j = 1,2,\cdots)$$

上式也称为 ξ 与 η 的联合分布。显然 $p_{ij} \geq 0$，$i,j = 1,2,\cdots$ 并且 $\sum_i \sum_j p_{ij} = 1$。

例如，我们可以写出两只股价变动的概率分布。令 $\xi_i = 1$ 表示股价出现上涨，$\xi_i = 0$ 表示股价出现下跌，其中 $i = 1$ 表示第一只股票，$i = 2$ 表示第二只股票，这样我们就可以写出 (ξ_1,ξ_2) 的分布如下②：

$$p(\xi_1 = 0,\xi_2 = 0) = 0.60$$

$$p(\xi_1 = 0,\xi_2 = 1) = 0.15$$

$$p(\xi_1 = 1,\xi_2 = 0) = 0.20$$

$$p(\xi_1 = 1,\xi_2 = 1) = 0.10$$

列成分布表如表 1－4 所示：

表 1－4

ξ_1 \ ξ_2	0	1
0	0.60	0.15
1	0.20	0.10

定义 1.2.7 **如果存在一个非负函数 $\varphi(x,y)$，使得二元随机变量 (ξ,η) 的分布函数 $F(x,y)$ 对于任意的实数 x,y 都有：$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y}\varphi(s,t)dsdt$。则称 (ξ,η) 是二元连续**

② 注：这里的数据均为假设，感兴趣的读者可以查阅计算出任意两只股票的概率分布。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



型随机变量。$\varphi(x,y)$ 称为 ξ 与 η 的联合分布密度。

$\varphi(x,y)$ 的性质：

（1）对一切实数 x,y，$\varphi(x,y) \geq 0$；（2）$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)dxdy = 1$。

显然，对任意实数 $a < b$ 及 $c < d$，有 $p(a < \xi \leq b, c < \eta \leq d) = \int_a^b\int_c^d\varphi(x,y)dxdy$。

三、独立性

（一）事件的独立性

定义 1.2.8 如果事件 A 发生的可能性不受事件 B 发生与否的影响，即 $P(A|B) = P(A)$，则称事件 A 对于事件 B 独立。显然，若 A 对于 B 独立，B 对于 A 也一定独立，我们称事件 A 与事件 B 相互独立。A 与 B 独立的充分必要条件是 $P(AB) = P(A) \cdot P(B)$。

（二）随机变量的相互独立性

定义 1.2.9 对于任何实数 x,y，如果二元随机变量 (ξ,η) 的联合分布函数 $F(x,y)$ 等于 ξ 与 η 的边缘分布函数的乘积，即 $F(x,y) = F_\xi(x) \cdot F_\eta(y)$，则称随机变量 ξ 与 η 相互独立。

定义 1.2.10 离散型二元随机变量 (ξ,η) 中，分量 ξ（或 η）的概率分布称为 (ξ,η) 的关于 ξ（或 η）的边缘分布。

四、随机变量函数的概念和分布

我们常常遇到一些随机变量，它们的分布往往难于直接得到，但是与它们有关系的另一些随机变量的分布却是容易知道的。因此，就要研究这两种随机变量之间的关系，然后通过它们之间的关系，由已知的随机变量的分布求出与之有关的其它随机变量的分布。

定义 1.2.11 设 $f(x)$ 是定义在随机变量 ξ 的一切可能取值集合上的函数。如果对于 ξ 的每一可能值 x，都有另一随机变量 η 的取值 $y = f(x)$ 与之相对应，则称 η 为 ξ 的函数，记作 $\eta = f(\xi)$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第三节　对总体的描述——随机变量的数字特征

通过前一节的分析，我们已经了解，总体的数据产生过程就是一个随机变量。因而，如何对总体进行描述的问题就是如何对随机变量进行描述的问题。在前一节中，主要讨论了随机分量的分布，这实质上就是对随机变量的一种最完全的描述。然而求出总体分布往往不是一件容易的事。同时，在很多情况下，我们并不需要全面地考察随机变量的变化情况，而只需了解它的一些综合指标就够了。一般来讲，我们常常需要了解随机变量的一般水平和它的偏离（离散）程度这两个指标。很像在进行投资分析时，我们需要经常考虑回报与风险这两个指标一样。了解了这两个指标后，就对随机变量及总体有了一个粗略的认识。在很多情况下，了解这两点也是进一步求取总体分布的基础和关键。所谓随机变量的数字特征，就是用数字来表示的随机变量的一些重要综合指标。

本节将讨论总体的两个最常用的数字特征：一是数学期望，它是用来描述随机变量及总体的一般水平的；二是方差，它是用来描述随机变量及总体的偏离程度的。在投资分析中，回报经常用均值（数学期望）来表示，风险程度经常用方差来表示。

一、数学期望（Expected Value）

定义 1.3.1　假定一个离散型随机变量 x 有 n 个不同的可能取值 $x_1,\cdots,x_n$，而 $p_1,\cdots,p_n$ 是它们相应的被取到的概率，则通过随机变量 x，数学期望 $E(x)$ 由下式定义：

$$E(x)=\mu_x=p_1x_1+p_2x_2+\cdots+p_nx_n=\sum_{i=1}^{n}p_ix_i$$

由上可见，$E(x)$ 实际上是随机变量 x 的所有可能取值 $x_1,\cdots,x_n$ 的加权平均数。

数学期望具有以下性质（x 为随机变量）：

1. 如果 a,b 为常数，则 $E(ax+b)=aE(x)+b$；
2. 如果 x,y 为两个随机变量，则 $E(x+y)=E(x)+E(y)$；
3. 如果 $g(x)$ 和 $f(x)$ 分别为 x 的两个函数，则

$$E[g(x)+f(x)]=E[g(x)]+E[f(x)]=\sum_{i=1}^{n}g(x_i)\cdot p_i+\sum_{i=1}^{n}f(x_i)\cdot p_i；$$

4. 如果 x,y 是两个独立的随机变量，则 $E(x\cdot y)=E(x)\cdot E(y)$。

注意： 数学期望是用来描述随机变量（或总体）的一般水平的。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

例如，我们可以使用表 1－2 中的数据，计算出该股票股价变化这一随机变量的期望值：

$$E\xi = p_1x_1 + p_2x_2 = 0.55 \times 0 + 0.45 \times 1 = 0.45$$

定义 1.3.2 **设连续型随机变量 x 有分布密度 $\varphi(x)$，若积分 $\int_{-\infty}^{\infty} x\varphi(x)dx$ 绝对收敛，则 $E(x) = \int_{-\infty}^{\infty} x\varphi(x)dx$ 称为 x 的数学期望。**

连续型随机变量的数学期望也满足上面的 4 个基本性质。

二、方差（Variance）

定义 1.3.3 **如果随机变量 x 的数学期望 $E(x)$ 存在，称 $[x-E(x)]$ 为随机变量 x 的离差。显然，随机变量离差的期望是 0，即 $E[x-E(x)] = 0$。**

定义 1.3.4 **随机变量离差平方的数学期望，叫做随机变量的方差，记作 $Var(x)$，$V(x)$ 或 $D(x)$，$\sqrt{Var(x)}$ 称为 x 的标准差（或方差根），即**

$$\sigma_x^2 = V(x) = Var(x) = E[x-E(x)]^2 = E(x-\mu_x)^2 。$$

如果 x 为连续型随机变量，则 $Var(x) = \int_{-\infty}^{\infty}[x-E(x)]^2\varphi(x)dx$。

如果 x 为离散型随机变量，则 $Var(x) = \sum_i E[x_i - E(x)]^2 \cdot p_i$。

注意： 1. 离差和方差都是用来描述随机变量 x 的离散程度的，亦即描述随机变量 x 相对于它的期望（平均值）的偏差程度，这种偏差越大，表明随机变量的取值越分散。

2. 在一般情况下，我们采用方差来描述离散程度。这是因为离差中包含有正偏差和负偏差，这使得在把离差加和时，正负相抵，和为 0，无法体现随机变量的总偏离程度。事实上，无论是正偏差大还是负偏差大，同样是离散程度大。方差中由于有平方，从而消除了正负号的影响，并易于加总，也易于强调大的偏离的突出作用。

方差具有以下性质：

1. $Var(c) = 0$（c 为常数）；
2. x 为随机变量，c 为常数，则 $Var(c+x) = Var(x)$；
3. x 为随机变量，c 为常数，则 $Var(cx) = c^2Var(x)$，$Var(-cx) = c^2Var(x)$；
4. x 为随机变量，a,b 为常数，则 $Var(a+bx) = b^2Var(x)$；

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

5. x,y 为两个相互独立的随机变量，则 $Var(x+y)=Var(x)+Var(y)=Var(x-y)$；

6. a,b 为常数，x,y 为两个相互独立的随机变量，则 $V(ax+by)=a^2Var(x)+b^2Var(y)$；

7. x 为随机变量，则 $Var(x)=E(x^2)-[E(x)]^2$。

我们同样可以使用表 1－2 中的数据，计算出该股票股价变化这一随机变量的方差：

$$\begin{aligned}Var(\xi) &= p_1(x_1-Ex)^2+p_2(x_2-Ex)^2\\ &= 0.55(0-0.45)^2+0.45(0.45-0.45)^2\\ &= 0.11\end{aligned}$$

三、数学期望与方差的图示

1. 期望（见图 1.3）

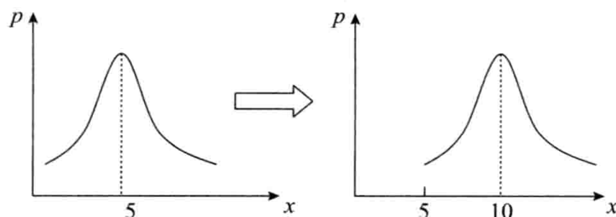


图 1.3 方差相同，期望变大

2. 方差（见图 1.4）

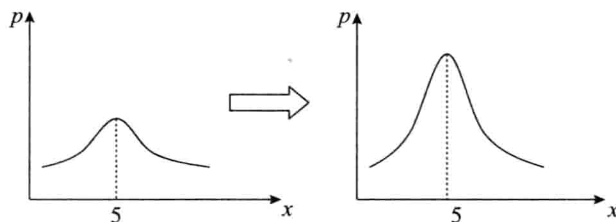


图 1.4 期望相同，方差变小

可见，方差描述随机变量的离散程度，而数学期望则描述随机变量的一般水平（平均水平）。

注意： 图 1.3、图 1.4 为示意性分析图。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第四节 对样本的描述——样本分布的数字特征

一、样本分布函数

前面已经讲过，总体的数据产生过程就是一个随机变量。因而，如果把随机变量 ξ 的分布看作某个总体的分布，那么 ξ 的分布函数 $F(x)$ 就是一个总体分布函数。

设 $(x_1,\cdots,x_n)$ 为总体 ξ 的一个样本观察值，把它们按大小排列为 $x_1^* \leq \cdots \leq x_n^*$，令

$$F_n(x) = \begin{cases} 0 & x < x_1^* \\ \frac{1}{n} & x_1^* \leq x < x_2^* \\ \cdots & \\ \frac{k}{n} & x_{n-1}^* \leq x < x_n^* \\ 1 & x_n^* \leq x \end{cases}$$

这里 $F_n(x)$ 等于样本的 n 个观察值中不超过 x 的个数除以样本容量 n，我们称它为样本分布函数。

二、样本平均数

定义 1.4.1 对于样本 $(x_1,\cdots,x_n)$，称 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 为样本平均数。

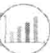
样本平均数是用来描述样本的一般平均水平。

三、样本方差

定义 1.4.2 对于样本 $(x_1,\cdots,x_n)$，称 $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 以及 $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ 分别为样本的方差和标准差。

注意：1. $\because \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \therefore s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$；2. 样本方差和标准差是用来描述样本的离散程度的。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第五节 随机变量的分布——总体和样本的连接点

本节将首先讨论几种重要的分布和它们的相互关系。然后，讨论总体和样本是怎样通过随机变量的分布被联系在一起的。在这一部分里，许多定理的证明被略去，只强调定理的结论本身。

一、几种重要的分布

鉴于本章的写作意图，我们在给出各种分布之前首先必须强调，请读者把注意力集中在各个分布之间的关系上，而不是某个分布的具体分布函数表达式上。这一点，是由经济计量学科的研究性质所决定的。因而，除 Γ 分布、χ^2 分布和正态分布外，其余各分布（F 分布、t 分布）的分布密度表达式我们也同样给出，有兴趣的读者可以参考。

这里有必要树立这样一种观念，如果一个随机变量的分布已经确定，那么这个随机变量的一切性质对于我们便都是已知的。换句话说，随机变量的分布是对随机变量最为完备的描述。举例来讲，如果已知随机变量 x 服从方差和期望已知（即分布函数已知）的正态分布，那么，我们便了解了这个随机变量的一切自身性质，并可以通过查阅正态分布表来确定在进一步研究中所需的一切临界值或其它有用的数据，而 x 的具体表达式、图形等等都只是技术性的问题。总之，我们的精力应该放在确定一个随机变量到底具有何种分布上。

1. Γ 分布

定义 1.5.1 **如果连续型随机变量 ξ 具有密度函数**

$$\varphi(x) = \begin{cases} \dfrac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & x > 0 (r > 0, \lambda > 0) \\ 0 & x \le 0 \end{cases}$$

则称 ξ 服从 Γ 分布，记作 $\Gamma(\lambda, r)$ 。

这里，$\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x} dx$ 。

当 $r > 0$ 时，$\Gamma(r)$ 积分收敛，且有 $\int_{-\infty}^{\infty} \varphi(x) dx = 1$ 。

定理 1.5.1 **对于根据定义 1.5.1 定义的 ξ，有 $E(\xi) = r/\lambda, V(\xi) = r/\lambda^2$**

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

2. 指数分布

定义 1.5.2 在定义 1.5.1 中，如果 $r = 1$，则我们称此时的 Γ 分布为指数分布密度函数。其表达式为

$$\varphi(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

定理 1.5.2 指数分布的期望为 $\frac{1}{\lambda}$，方差为 $\frac{1}{\lambda^2}$。

3. χ^2 分布

定义 1.5.3 $r = \frac{n}{2}$（n 为正整数），$\lambda = \frac{1}{2}$ 的 Γ 分布称为具有 n 个自由度的 χ^2 分布，记为 $\chi^2(n)$。其密度函数为

$$\varphi(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0 \\ 0, & x \leq 0 \end{cases}$$

定理 1.5.3 若 $\xi_1, \cdots, \xi_n$ 相互独立，且 ξ_i 服从参数为 (λ, r_i) 的 Γ 分布 $(i = 1, \cdots, n)$，则它们的和 $\xi_1 + \cdots + \xi_n$ 服从参数为 $(\lambda, r_1 + \cdots + r_n)$ 的 Γ 分布。

推论 若 $\xi_1, \cdots, \xi_n$ 相互独立，且 ξ_i 服从具有 $n_i (i = 1, 2, \cdots, n)$ 个自由度的 χ^2 分布，则它们的和 $\xi_1 + \cdots + \xi_n$ 服从具有 $\sum_{i=1}^{k} n_i$ 个自由度的 χ^2 分布。

4. 正态分布

定义 1.5.4 若连续型随机变量 ξ 的概率密度为 $\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$（$\sigma$，$\mu$ 为常数，$\sigma > 0$），则称 ξ 服从正态分布，简记 $\xi \sim N(\mu, \sigma^2)$。

定理 1.5.4 正态分布的期望 $E(\xi) = \mu$，方差 $Var(\xi) = \sigma^2$。

可见，正态分布概率密度中的两个参数 μ 和 σ，分别是随机变量的期望和标准差。因而，已知期望和方差，可以完全确定一个正态分布。

定义 1.5.5 当 $\mu = 0$，$\sigma = 1$ 时的正态分布，称为标准正态分布，记作 $\xi \sim N(0,1)$。其 $\varphi(x)$ 为 $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



定理 1.5.5 如果 $\xi \sim N(\mu,\sigma^2)$，且 $\eta = \dfrac{\xi-\mu}{\sigma}$，那么 $\eta \sim N(0,1)$。

利用定理 1.5.5，可以将任何一个正态分布，化为标准正态分布，即将其标准化。

5. t 分布

定义 1.5.6 若连续型随机变量 ξ 的分布密度 $\varphi(x)$ 由下式给出，则称 ξ 服从具有 n 个自由度的 t 分布，记作 $t(n)$：

$$\varphi(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\cdot\sqrt{n\pi}}(1+\frac{x^2}{2})^{-\frac{n+1}{2}}$$

6. F 分布

定义 1.5.7 若连续型随机变量 ξ 的分布密度 $\varphi(x)$ 由下式给出，则称 ξ 服从第一个自由度为 n_1，第二个自由度为 n_2 的 F 分布，记作 $F(n_1,n_2)$。

$$\varphi(x) = \begin{cases} \dfrac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\cdot\Gamma(\frac{n_2}{2})}(\dfrac{n_1}{n_2})\cdot x^{\frac{n_1}{2}-1}(1+x)-\dfrac{n_1+n_2}{2} & x>0 \\ 0 & x\leq 0 \end{cases}$$

7. 各分布的图象

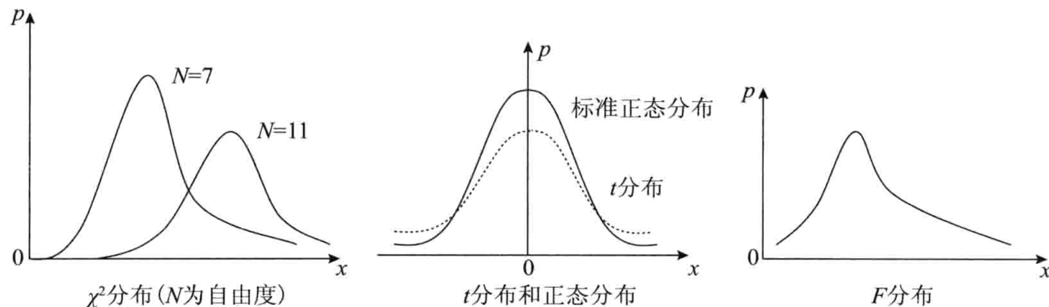


图 1.5

注意：1. t 分布和标准正态分布均关于 y 轴对称，即：$\varphi_0(-x) = \varphi_0(x)$；2. t 分布和标准正态分布均在 $x=0$ 处取得最大值。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

二、各分布之间的关系

（一）三大分布的有关性质

1. χ^2 分布的性质

（1）可加性。若 $x \sim \chi^2(m)$，$y \sim \chi^2(n)$ 且 x 与 y 相互独立，则 $(x+y) \sim \chi^2(m+n)$；

（2）若 $x \sim \chi^2(m)$，则 $Ex = m, Dx = 2m$。

2. F 分布的性质

（1）随机变量 F 恒为正值，F 分布也是一个连续的非对称分布；

（2）分布具有一定程度的反对称性。

3. t 分布的性质

自由度很大的时候（大于 45），t 分布接近标准正态分布。

（二）各分布之间的关系

1. 一般正态分布与标准正态分布之间的关系

定理 1.5.6 如果 $\xi \sim N(\mu,\sigma^2)$，则 $\frac{\xi-\mu}{\sigma} \sim N(0,1)$。

2. 标准正态分布与 χ^2 分布之间的关系

定理 1.5.7 如果 $\xi \sim N(0,1)$，则 $\xi^2 \sim \chi^2(1)$，即 ξ^2 服从具有一个自由度的 χ^2 分布。

3. 标准正态分布与 t 分布之间的关系

定理 1.5.8 设两个随机变量 ξ 与 η 相互独立，且 $\xi \sim N(0,1)$，$\eta \sim \chi^2(n)$，则 $T = \frac{\xi}{\sqrt{\eta/n}}$ 服从具有 n 个自由度的 t 分布，即 $T \sim t(n)$，其密度函数见定义 1.5.6。

4. 标准正态分布与 F 分布之间的关系（或说 χ^2 分布与 F 分布的关系）

定理 1.5.9 设两个随机变量 ξ_1 和 ξ_2 相互独立，且 $\xi_1 \sim \chi^2(n_1)$，$\xi_2 \sim \chi^2(n_2)$，则有 $F = \frac{\xi_1/n_1}{\xi_2/n_2} \sim F(n_1,n_2)$。其中 $F(n_1,n_2)$ 表示第一个自由度为 n_1，第二个自由度为 n_2 的 F 分布，其密度函数见定义 1.5.7。

5. 关于正态分布的和

定理 1.5.10 设 $x_1,\cdots,x_n$ 相互独立，x_i 服从正态分布 $N(\mu_i,\sigma^2)$，则它们的线性函数 $\eta = \sum_{i=1}^{n} a_i x_i$（$a$ 不全为 0），也服从正态分布，且 $E\eta = \sum_{i=1}^{n} a_i\mu_i$，$Var(\eta) = \sum_{i=1}^{n} a_i^2\sigma_i^2$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

6. 关于χ^2 分布（参阅定理 1.5.3 推论）

定理 1.5.11 设 $x_1,\cdots,x_n$ 相互独立，都服从标准正态分布，则 $x^2 = \sum_{i=1}^{n} x_i^2$ 服从具有 n 个自由度的χ^2 分布，记为$\chi^2(n)$。

定理 1.5.12 设 $x_1,\cdots,x_n$ 相互独立，都服从标准正态分布，则它们的平均数 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 与对平均数 $\bar{x}$ 的离差平方和 $\sum_{i=1}^{n}(x_i - \bar{x})^2$ 相互独立，且 $\sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \chi^2(n-1)$。

为了便于记忆上述 7 个定理，读者可以这样设想，如果 $X \sim N(0,1)$，则定理 1.5.7 表明 $X^2 \sim \chi^2$，定理 1.5.8 表明 $\frac{X}{\sqrt{X^2}} \sim t$，定理 1.5.9 表明 $\frac{X^2}{X^2} \sim F$，定理 1.5.10 表明 $\sum X \sim N$，定理 1.5.11 表明 $\sum X^2 \sim \chi^2$。

三、分布是样本与总体的连接点

定理 1.5.13 设 $x_1,\cdots,x_n$ 是取自正态总体 $N(\mu,\sigma^2)$ 的样本，则有：

$$\bar{x} \sim N(\mu,\frac{\sigma^2}{n})\ ;\ \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

证明： 定理 1.5.10 中取 $a_i = \frac{1}{n}(i = 1,\cdots,n)$，有 $\bar{x} \sim N(\mu,\frac{\sigma^2}{n})$，由此结合定理 1.5.6，可得到以上结论。

定理 1.5.14 设 $x_1,\cdots,x_n$ 是取自正态总体 $N(\mu,\sigma^2)$ 的样本，则有 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \chi^2(n-1)$，$\bar{x}$ 与 $\sum_{i=1}^{n}(x_i - \bar{x})^2$ 相互独立。

证明： $\because$ $x_1,\cdots,x_n$ 均取自正态总体 $N(\mu,\sigma^2)$

$$\therefore \frac{x_i-\mu}{\sigma} \sim N(0,1) \qquad (i = 1,2,\cdots,n，根据定理 1.5.6)$$

据定理 1.5.12，$\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma} - \frac{\bar{x}-\mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \chi^2(n-1)$

定理 1.5.15 设 $x_1,\cdots,x_n$ 是取自正态总体 $N(\mu,\sigma^2)$ 的样本，$\bar{x}$，s 分别是样本的平均数和标准差，则 $T = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t(n-1)$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

证明：根据定理 1.5.13，得 $\xi = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

根据定理 1.5.14，得 $\eta = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2} \sim \chi^2(n-1)$

根据定理 1.5.8，得 $T = \dfrac{\xi}{\eta/\sqrt{n-1}} \sim t(n-1)$

$$\text{而 } T = \frac{\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}}{n-1}}} = \frac{\sqrt{n}(\bar{x}-\mu)}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}} = \frac{\sqrt{n}(\bar{x}-\mu)}{s} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

其中，$s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$，故 $\dfrac{\bar{x}-\mu}{s/\sqrt{n}} \sim t(n-1)$。

定理 1.5.16 设 $x_1,\cdots,x_{n1}$ 和 $y_1,\cdots,y_{n2}$ 分别是来自两个相互独立的正态总体 $N(\mu_1,\sigma^2)$ 和 $N(\mu_2,\sigma^2)$ 的样本，则

$$T = \frac{\bar{x}-\bar{y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\cdot\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2)$$

其中，$\bar{x}$、$\bar{y}$、s_1^2、s_2^2 分别为两个样本各自的平均数和方差。

定理 1.5.17 设 $x_1,\cdots,x_{n1}$ 和 $y_1,\cdots,y_{n2}$ 分别是来自两个相互独立的正态总体 $N(\mu_1,\sigma_1^2)$ 和 $N(\mu_2,\sigma_2^2)$ 的样本，则

$$F = \frac{1}{\sigma_1^2}s_1^2 \Big/ \frac{1}{\sigma_2^2}s_2^2 \sim F(n_1-1,n_2-1)$$

其中，s_1^2、s_2^2 分别为两个样本各自的方差。

证明：根据定理 1.5.14，

$$\xi_1 = \frac{1}{\sigma_1^2}\sum_{i=1}^{n_1}(x_i-\bar{x}) \sim \chi^2(n_1-1),\ \xi_2 = \frac{1}{\sigma_2^2}\sum_{i=1}^{n_2}(y_i-\bar{y}) \sim \chi^2(n_2-1)$$

根据定理 1.5.9，得 $T = \dfrac{\xi_1/n_1-1}{\xi_2/n_2-1} \sim F(n_1-1,n_2-1)$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\text{而 } T=\frac{\frac{1}{\sigma_1^2}\sum_{i=1}^{n_1}(x_i-\bar{x})^2/(n_1-1)}{\frac{1}{\sigma_2^2}\sum_{i=1}^{n_2}(y_i-\bar{y})^2/(n_2-1)}=\frac{\frac{1}{\sigma_1^2}s_1^2}{\frac{1}{\sigma_2^2}s_2^2}$$

$$\therefore\ \frac{1}{\sigma_1^2}s_1^2\Big/\frac{1}{\sigma_2^2}s_2^2\sim F(n_1-1,n_2-1)$$

注意：1. 定理 1.5.13、1.5.14、1.5.15、1.5.16、1.5.17 可分别视为定理 1.5.10、1.5.12、1.5.7、1.5.8、1.5.9 的推论。

2. 定理 1.5.13 ~ 1.5.17 的实质是总体与样本相联系的纽带。在本章第一节中，我们已知道，所谓总体就是一个随机变量，所谓样本就是 n 个相互独立且与总体有相同分布的随机变量 $x_1,\cdots,x_n$，即一个 n 元随机变量 $(x_1,\cdots,x_n)$。这里，应注意，总体与样本之间的联系在于具有相同的分布。而本节的任务就是使这一点具体化，从而为进一步通过样本的特征来估计和代替总体的特征（这是我们研究的根本任务和方向）铺平道路。对此，我们以定理 1.5.15 为例作如下分析：

设想，当我们研究某一个问题时，要利用某一个随机变量 $\xi\sim N(\mu,\sigma^2)$ 这一条件，我们根据定理 1.5.6，常常首先把 ξ 标准化，即 $\frac{\xi-\mu}{\sigma}\sim N(0,1)$，然后查标准正态分布表，获得我们所需要的信息，如临界值等。然而，假设我们对 ξ 的分布了解得并不完备，我们只知道它属于某一个期望为 μ 的正态分布，而不知其方差 σ^2 的具体值。在这种情况下，显然无法再利用 $\frac{\xi-\mu}{\sigma}\sim N(0,1)$ 这一关系式。很自然，我们想到了一种解决困难的办法，即随机地从该随机变量所刻画的母体中抽取一组样本，并计算这组样本的方差 s^2，进而用 s^2 代替未知的总体真实方差 σ^2 来估算 $\frac{\xi-\mu}{\sigma}$。这里，我们就是在用样本的数量特征 s^2 来代替总体的数量特征 σ^2 进行估计了，而这样对 $\frac{\xi-\mu}{\sigma}$ 估计所得的结果 $\frac{\xi-\mu}{\sigma}$ 是否仍服从 $N(0,1)$ 呢？如果不是，它服从什么分布呢？对此，定理 1.5.15 给出了答案，即 $\frac{\xi-\mu}{\sigma}\sim t(n-1)$。

3. 第 2 点中所表达的思想方法将在以后的区间估计和假设检验两部分中得到广泛应用。这两部分内容，多次使用了本节中所给出的定理和描述的思想方法。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第六节　通过样本，估计总体（一）——估计量的特征

从这一节开始，我们研究如何通过样本特征来估计总体特征的问题。首先，讨论估计量的特征。

所谓估计量的特性，就是这样一些特性，当估计量满足了这些特性时，我们说该估计量是一个相对较好、较合理的估计量，否则，就是一个不够完善的估计量。可见，估计量的特性实际上是衡量一估计量的好坏标准。具备这些特性，是对一个估计量的基本要求。对于总体的某一个数字特征，不同人可以提出各种不同的估计量，在这些估计量中，只有那些满足若干上述特性的才可以被接受。这里我们举出四个这样的特性：

一、无偏性（Lack of Bias）

根据样本推得的估计值和真值可能不同，然而，如果有一系列抽样构成各个估计（这些估计构成一个随机变量），很自然的会要求这些估计的期望值与未知参数的真值相等。这一点的直观意义是：样本估计量的数值在真值周围摆动，而无系统误差。

定义 1.6.1　如果 $E\hat{\theta}=\theta$ 成立，我们称 $\hat{\theta}$ 为参数 θ 的无偏估计，亦称 $\hat{\theta}$ 具有无偏性。

如果 $E\hat{\theta}\neq\theta$，我们称 $\hat{\theta}$ 为 θ 的有偏估计，其偏差为 $\text{Bias}=E(\hat{\theta})-\theta$。如图 1.6 所示：

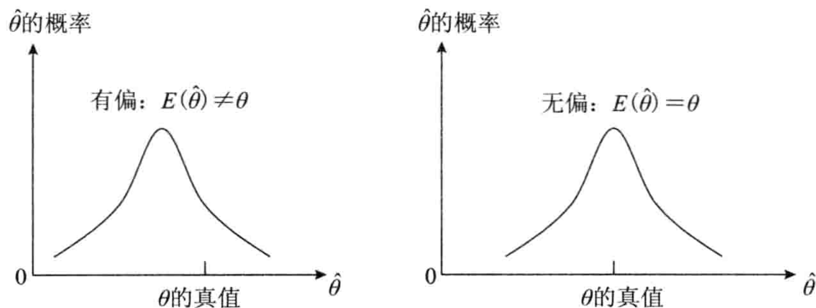


图 1.6

前面我们曾指出，样本平均数和样本方差（$\bar{x}$ 和 s^2）常常被用作总体的 $E(x)$ 和 σ^2 的估计量，而且 s^2 定义本身就含有使 s^2 成为 σ^2 的无偏估计的意图。这里我们给出证明，从总体 ξ 中取一组样本 $(x_1,\cdots,x_n)$，$E\xi=\mu$，$V\xi=\sigma^2$，其样本平均数 $\bar{x}$ 和样本方差 s^2 分别

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



是 μ 及 σ^2 的无偏估计。

证明： $E\bar{x} = E(\frac{1}{n}\sum_{i=1}^{n}x_i) = \frac{1}{n}\sum_{i=1}^{n}Ex_i = \frac{1}{n}\cdot n\cdot\mu = \mu$

$$Var(\bar{x}) = V(\frac{1}{n}\sum_{i=1}^{n}x_i) = \frac{1}{n^2}\sum_{i=1}^{n}Var(x_i) = \frac{1}{n^2}\cdot n\cdot\sigma^2 = \frac{\sigma^2}{n}$$

$$E(s^2) = E[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})]^2$$

$$= E\left\{\frac{1}{n-1}\sum_{i=1}^{n}[x_i-\mu-(\bar{x}-\mu)]^2\right\}$$

$$= \frac{1}{n-1}E[\sum_{i=1}^{n}(x_i-\mu)^2 - 2\sum_{i=1}^{n}(x_i-\mu)(\bar{x}-\mu) + \sum_{i=1}^{n}(\bar{x}-\mu)^2]$$

$$= \frac{1}{n-1}E[\sum_{i=1}^{n}(x_i-\mu)^2 - 2n(\bar{x}-\mu)^2 + n(\bar{x}-\mu)^2]$$

$$= \frac{1}{n-1}E\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{n}{n-1}E(\bar{x}-\mu)^2$$

$$= \frac{1}{n-1}\cdot n\cdot\sigma^2 - \frac{n}{n-1}\cdot\frac{\sigma^2}{n} = \sigma^2$$

这里 $E(x_i-\mu)^2 = \sigma^2$，$E(\bar{x}-\mu)^2 = Var(\bar{x})$

因此，$E(\bar{x}) = \mu$，$E(s^2) = \sigma^2$，$\bar{x}$ 和 s^2 分别是 μ 和 σ^2 的无偏估计量。

注意：

1. 如果从总体中随机取出两个相互独立的样本，其容量分别为 n_1 和 n_2，则可以证明 $\bar{x} = \frac{1}{n_1+n_2}(n_1\bar{x}_1+n_2\bar{x}_2)$，$s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ 分别是总体中 μ 和 σ^2 的无偏估计量，其中 $s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n}(x_{ij}-\bar{s}_i)^2(i=1,2)$。

2. 无偏性是对一个估计量最重要的要求之一，但它只能保证估计量的期望等于真值。而且，对于总体的某个特定参数，其无偏估计量往往不只一个。

例如，可以验证 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$ 和 $x' = \frac{1}{\sum_{i=1}^{n}a_i}\sum_{i=1}^{n}a_ix_i(\sum_{i=1}^{n}a_i \neq 0)$ 都是总体期望值 μ 的无偏估计量。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



3. 由前面证明可知，$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ 不是方差 σ^2 的无偏估计量，这一点，我们以后还会提到。

二、有效性（Efficiency）

我们已经知道，总体的某一个参数 θ 的无偏估计量往往不只一个，而且无偏性仅仅表明 $\hat{\theta}$ 所有可能取的值按概率平均等于 θ，它取的值可能大部分与 θ 相差很大。为保证 $\hat{\theta}$ 的取值能集中于 θ 附近，必须要求 $\hat{\theta}$ 的方差越小越好。

定义 1.6.2 **设 $\hat{\theta}$ 和 $\hat{\theta}'$ 都是 θ 的无偏估计，若对任意的样本容量 n，有 $\hat{\theta}$ 的方差总小于 $\hat{\theta}'$ 的方差，则称 $\hat{\theta}$ 是比 $\hat{\theta}'$ 有效的估计量。如果在 θ 的一切无偏估计量中，$\hat{\theta}$ 的方差达到最小，则 $\hat{\theta}$ 称为 θ 的有效估计量，亦称 $\hat{\theta}$ 具有有效性（见图 1.7）。**

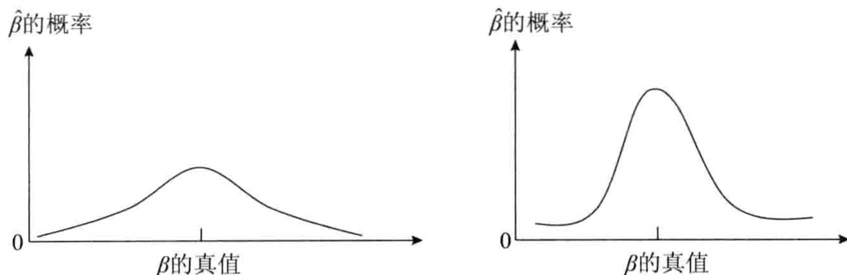


图 1.7

比较总体期望值 μ 的两个无偏估计 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 与 $x'=\frac{1}{\sum_{i=1}^{n}a_i}\sum_{i=1}^{n}a_ix_i\left(\sum_{i=1}^{n}a_i\neq 0\right)$ 的有效性。

首先，可以证明它们的无偏性：

$$E\bar{x}=\mu$$

$$Ex'=\frac{1}{\sum_{i=1}^{n}a_i}\sum_{i=1}^{n}a_iEx=\mu$$

下面，证明它们的有效性：

$$V(\bar{x})=\frac{1}{n}\sigma^2$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$V(x')=\frac{1}{(\sum_{i=1}^{n}a_i)^2}\sum_{i=1}^{n}a_i^2V(x_i)=\frac{\sum_{i=1}^{n}a_i^2}{(\sum_{i=1}^{n}a_i)^2}\sigma^2$$

利用不等式 $a_i^2+a_j^2\geq 2a_ia_j$

$$(\sum_{i=1}^{n}a_i)^2=\sum_{i=1}^{n}a_i^2+\sum_{i<j}^{n}2a_ia_j\leq\sum_{i=1}^{n}a_i^2+\sum_{i<j}^{n}(a_i^2+a_j^2)=n\sum_{i=1}^{n}a_i^2$$

$$V(x')\geq\frac{\sum_{i=1}^{n}a_i^2}{n\sum_{i=1}^{n}a_i^2}\sigma^2=\frac{1}{n}\sigma^2=V(\bar{x})$$

这一结果说明 $\bar{x}$ 比 x' 更有效。

注意：

1. 一个无偏有效估计量取的值是在可能范围内最密集于 θ 附近的。换而言之，它以最大的概率，保证该估计的观察值在未知参数的真值 θ 附近摆动。

2. 我们进一步可以证明，样本平均数 $\bar{x}$ 是总体期望值 $E(x)$ 的有效估计量。

三、最小离差平方性（Mean Square Error）

在很多情况下，我们被迫在偏差的大小和方差的大小之间作出抉择。比如，当我们以预测的确定性作为建立模型的目标时，一个方差极小的有偏估计比一个方差很大的无偏估计可能更为我们所追求。在这一方面，估计量的最小离差平方性为偏差与方差二者的权衡提供了一个有效的尺度。

定义 1.6.3 **若参数 θ 的估计量为 $\hat{\theta}$，$E(\hat{\theta}-\theta)^2$ 称为估计量 θ 的离差平方，记为 $MSE(\hat{\theta})$。在 θ 的一切估计量中，使其估计量离差平方最小的估计量，称为 θ 的最小离差平方估计量，具有最小离差平方性。**

我们可以证明：$MSE(\hat{\theta})=[Bias(\hat{\theta})]^2+Var(\hat{\theta})$

$$\begin{aligned}MSE(\hat{\theta})&=E(\hat{\theta}-\theta)^2\\&=E[(\hat{\theta}-E(\hat{\theta}))+(E(\hat{\theta})-\theta)]^2\\&=E[\hat{\theta}-E(\hat{\theta})]^2+E[E(\hat{\theta})-\theta]^2+2E[\hat{\theta}-E(\hat{\theta})][E(\hat{\theta})-\theta]\\&=Var(\hat{\theta})+[Bias(\hat{\theta})]^2+0\end{aligned}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\because 2E[\hat{\theta}-E(\hat{\theta})][E(\hat{\theta})-\theta]=2[E(\hat{\theta})-\theta]\cdot E[\hat{\theta}-E(\hat{\theta})]=0$$

注意：

1. 由上可见，$MSE(\hat{\theta})$ 是 $\hat{\theta}$ 的偏差平方和 $\hat{\theta}$ 的方差之和。使 $MSE(\hat{\theta})$ 最小（即求 Min $MSE(\hat{\theta})$），就是使 $\hat{\theta}$ 偏差的平方与 θ 的方差之和为最小。因而，$MSE(\hat{\theta})$ 这一指标，考虑了偏差和方差两方面的因素，并给出了一个对这两方面进行权衡的方法（见图 1.8）。

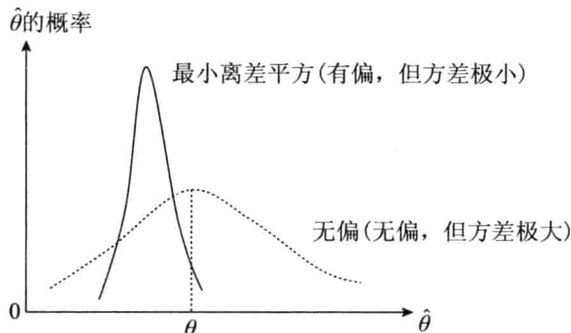


图 1.8

2. 若本身为无偏估计量（即 $Bias(\hat{\theta})=0$），那么 $\hat{\theta}$ 的离差平方 MSE 就等于 $\hat{\theta}$ 的方差，$MSE(\hat{\theta})=Var(\hat{\theta})+[Bias(\hat{\theta})]^2=Var(\hat{\theta})+0=Var(\hat{\theta})$。在此情况下，一个估计量如果具有最小离差平方性，那么它一定具有有效性，即 Min $MSE(\hat{\theta})$ = Min $Var(\hat{\theta})$。换言之，无偏估计量的最小离差平方性等价于它的有效性。

四、一致性

我们知道在一般情况下 $\hat{\theta}\neq\theta$，但我们有理由希望当样本容量 n 接近无限大时，$\hat{\theta}$ 依概率收敛于 θ。换言之，尽管根据样本求得的未知参数 θ 的估计量 $\hat{\theta}$ 常与这个参数的真值不同，我们仍希望当 $n\to\infty$时，估计量 $\hat{\theta}$ 在其真值 θ 附近的概率趋近于 1。

定义 1.6.4 **（“依概率收敛”的定义）若存在常数 a，使对于任何 $\varepsilon>0$，有 $\lim\limits_{n\to\infty}P(|\xi_n-a|<\varepsilon)=1$，则称随机变量序列 $\{\xi_n\}$ 依概率收敛于 a。**

定义 1.6.5 **若当 $n\to\infty$时，θ 依概率收敛于 $\hat{\theta}$，即任给 $\varepsilon>0$，若 $\lim\limits_{n\to\infty}P(|\hat{\theta}-\theta|<\varepsilon)=1$，则称 $\hat{\theta}$ 为参数 θ 的一致估计量，具有一致性（见图 1.9）。**

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



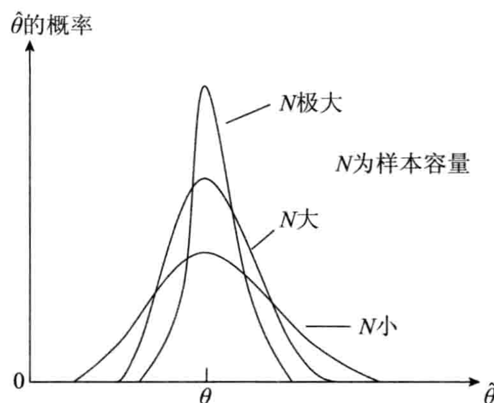


图 1.9

注意：

1. 一致性是从极限性质来看的，它只在样本容量较大时才起作用。

2. 作为评价估计量好坏的一个标准，计量经济学家在估计量的一致性和无偏性二者中更偏重于选择一致性。

3. 如果一个有偏的估计量具备一致性，那么虽然它在平均水平上可能与参数的真值不同，但根据定义，当样本信息量加大时，它会变得十分接近真值。因而，我们可以说一个有偏的一致性估计量具有大样本下的无偏性。同时，根据数理统计学中的大数定理，当样本容量变得越来越大时，从样本中获得的总体估计量具有相当的稳定性，即，其方差变得很小。因而，可以认为，一个有偏的一致性估计量在大样本下，同时具有“有效性”和“无偏性”。显然，它要比一个方差很大的无偏估计量优越得多。

4. 由于 $MSE(\hat{\beta}) = Var(\hat{\beta}) + [Bias(\hat{\beta})]^2$，估计量的一致性实质上等价于当 $n \to \infty$时，$MSE(\hat{\beta}) \to 0$。这是因为如果 $n \to \infty$，$MSE(\hat{\beta}) \to 0$，那么，$Var(\hat{\beta}) + [Bias(\hat{\beta})]^2 \to 0$。从而 $Var(\hat{\beta}) \to 0$，$[Bias(\hat{\beta})]^2 \to 0$，即随着样本加大，$\hat{\beta}$ 的方差变得很小，偏差接近于0，这正是一致性所描述的情况。事实上，一致性和 $MSE(\hat{\beta}) \to 0$（当 $n \to \infty$时）这两条标准在计量中往往是通用的。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第七节　通过样本，估计总体（二）——估计方法

估计总体参数的方法通常有两种：一是点估法，二是区间估计法。

一、点估法（Point Estimation）

所谓点估计就是给出被估计参数的一个特定的估计值。常用的点估法有四种，即矩法、最大似然估计法、最小二乘估计法和最小χ^2法。这四种方法是分别建立在不同的估计原则之上的。对同一个样本，根据这四种方法分别估计同一个参数，所获得的估计结果很可能互不相同。然而，由于各种原则都具备一定的合理性，所以以上四种估计方法都经常在研究中得到应用。下面，我们依次介绍矩法、最大似然法和最小二乘法三种方法。最小χ^2法由于在本书以后的章节中使用较少，故而略去。

1. 矩法（Moment Estimation Method）

矩法是求估计量最古老的方法。具体的作法是：以样本矩作为相应的总体矩的估计量，以样本矩的函数作为相应的总体矩同样函数的估计量。

这一方法最常见的应用是用样本平均数$\bar{x}$估计总体期望值μ，即认为$\hat{\mu}=\bar{x}$。

下面，我们给出一个纯数字的例子：

例　若样本$x_1, x_2, \cdots, x_n$取自均匀分布$\varphi(x,\theta)=\begin{cases}\frac{1}{\theta} & 0<x<\theta \\ 0 & 其它\end{cases}$，在矩法下，$\theta$是多少？

解　$\because \mu=\int_{-\infty}^{+\infty} x\varphi(x,\theta)dx=\int_0^{\theta} x\frac{1}{\theta}dx=\frac{1}{\theta}[\int_0^{\theta} xdx]=\frac{1}{\theta}[\frac{1}{2}x^2|_0^{\theta}]=\frac{\theta}{2}$

又$\because$在矩法下$\hat{\mu}=\bar{x}$ $\therefore \frac{\theta}{2}=\bar{x}$，$\hat{\theta}=2\bar{x}$

矩法比较直观，求估计量有时也比较直接，但它求出的估计量往往不够理想。

2. 最大似然估计法（Maximum—Likelihood Estimation Method）

我们首先来讨论最大似然的概念。与这一概念具有重要联系的是如下这一事实：不同的统计总体会产生出不同的样本，对于某一特定的样本，在我们这些不了解产生它的母体

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



究竟为何物的观察者眼中，它来自一些形式母体的可能性要比来自另一些的可能性大，即一些母体比另一些母体更容易产生出我们所观察到的样本。举例来说，假定我们抽取到了一个样本 $(x_1,x_2,\cdots,x_8)$，我们知道这一样本来自一个正态总体，这个正态总体的方差我们是了解的，但我们却不知它的期望是多少，如图 1.10 所示。假定这 8 个样本观察值不是来自 A 分布就是来自 B 分布，那么很明显，如果产生样本的真正分布是 B，那么我们观察到 $x_1,x_2,\cdots,x_8$ 这 8 个样本点的概率是非常小的。相反，如果真正的母体是 A，那么我们获得上述样本的可能性会显著增大。很显然，我们愿意承认 A 为真实的母体，因为 A 比 B 更可能产生出我们获得的样本观察值。在某种意义上，是样本“替”我们“选择”了总体 A。

假定现在要根据从总体 ξ 中抽取到的样本 $(x_1,\cdots,x_n)$，对总体分布中的未知数 θ 进行估计。最大似然法是选择这样的估计量 $\hat{\theta}$ 作为 θ 的估计值，使观察结果即样本 $(x_1,\cdots,x_n)$ 出现的可能性（概率）最大。对于离散型随机变量，我们就是要选择 $\hat{\theta}$ 使 $P(x_1)P(x_2)\cdots P(x_n)$ 最大；对于连续型随机变量，我们就是要选择 $\hat{\theta}$ 使 $\varphi(x_1)\varphi(x_2)\cdots\varphi(x_n)$ 最大。对于连续型随机变量，$\varphi(x_i)$ 表明随机变量在 x_i 附近取值的概率大小，因而相当于离散型随机变量中的 $P(x_i)$。

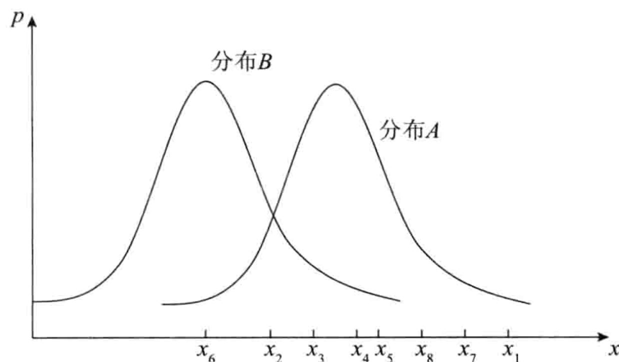


图 1.10

下面，我们用数学语言对上述思想进行说明。

设 ξ 为连续型随机变量，它的分布函数是 $F(x;\theta)$，分布密度是 $\varphi(x;\theta)$，其中 θ 是未知参数。由于样本的独立性，则样本 $(x_1,\cdots,x_n)$ 的联合分布密度是：

$$L(x_1,\cdots,x_n;\theta) = \prod_{i=1}^{n}\varphi(x_i;\theta)$$

由于每一取定的样本值 $x_1,\cdots,x_n$ 是常数，所以 L 可看成参数 θ 的函数。我们把 L 称为

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

样本的似然函数。又设 ξ 为离散型随机变量，有概率函数 $P(\xi = x_i) = P(x_i;\theta)$，则似然函数 $L(x_1,\cdots,x_n;\theta) = \prod_{i=1}^{n} P(x_i;\theta)$。

定义 1.7.1　如果 $L(x_1,x_2,\cdots,x_n;\theta)$ 在 $\hat{\theta}$ 处达到最大值，则称 $\hat{\theta}$ 是 θ 的最大似然估计。

为了取得 θ 的最大似然估计，我们必须使 L 达到最大值，并且把此时的 $\hat{\theta}$ 作为 θ 的估计量。由于 L 与 $\ln L$ 同时达到最大值，所以我们只需求 $\ln L$ 的最大值点即可，这样往往会给计算带来极大方便。

由于 L（即 $\ln L$）是参数 θ 的函数，根据微积分中的拉格朗日定理，$\ln L$ 的最大值应在 $\ln L$ 对 θ 的一阶导数等于0时取到。因而，考虑方程 $\frac{d\ln L}{d\theta} = 0$ 容易看出，我们所要 $\hat{\theta}$ 就是这个方程中 θ 的解。这个方程称为似然方程。

例　已知随机变量 ξ 的分布为：$\xi \sim \varphi(x_i,\theta) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}} & x > 0\ (\theta > 0) \\ 0 & \text{其它} \end{cases}$，$x_1,x_2,\cdots,x_n$ 是 ξ 的一组观察值，求 θ 的最大似然估计。

解　构造似然函数 $L(x_1,\cdots,x_n;\theta) = \prod_{i=1}^{n} \frac{1}{\theta}e^{-\frac{x_i}{\theta}} = \frac{1}{\theta^n} \cdot e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i}$

$$\ln L = -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

$$\frac{d\ln L}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$$

解似然方程：$\frac{d\ln L}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0$ 得 $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$

$\therefore$ $\bar{x}$ 是 θ 的最大似然估计。

例　已知 ξ 服从正态分布 $N(\mu,\sigma^2)$，$x_1,x_2,\cdots,x_n$ 为 ξ 的一组样本观察值，用最大似然法估计 μ,σ^2 的值。

解　$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2}}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \cdot \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$

$$\ln L = \frac{n}{2}\ln\left(\frac{1}{2\pi}\right) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

注意： 由于有两个参数，这里应分别将 L 对 μ 和 σ^2 求偏导数：

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$$

这里要解如下的似然方程组：

$$\begin{cases} \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i - \mu) = 0 \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \end{cases}$$

解得：

$$\begin{cases} \hat{\mu} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i = \bar{x} \\ \hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{cases}$$

注意：

1. 由本例我们知道，当不仅仅有一个总体分布参数需要估计时，应将 L 分别对各个不同的参数求偏导，然后解一个似然方程组。

2. 用最大似然法求出的总体方差的估计量 $\hat{\sigma}^2$ 是 $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 不是 σ^2 的无偏估计量，在本章第六节中讲过 $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 不是 σ^2 的无偏估计量，$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 才是 σ^2 的无偏估计。事实上，用最大似然法对方差进行估计所得到的估计量往往都是有偏的。这一点，会在以后用最大似然法对线性回归模型进行估计时再次体会到。

以上两个结论非常重要，请读者们谨记。

---

3. 最小二乘法

这里对最小二乘法暂不加以讨论，留待后面章节详述。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



二、区间估计

所谓区间估计，就是给出被估计参数所在的一个可能取值范围。

用点估计来估计总体参数，即使是无偏有效的估计量，也会由于样本的随机性，使得这些从样本算出的估计量的值并不恰是参数的真值。而且，即使估计值与真值真正相等，由于参数真值本身是未知的，我们也无从肯定这种相等。那么到底二者相差多少呢？这个问题换一个提法就是：根据估计量的分布，在一定的可靠程度下，指出被估计的总体参数所在的可能数值范围。这就是参数的区间估计问题。其具体做法是找两个统计量 $\theta_1(x_1,\cdots,x_n)$ 与 $\theta_2(x_1,\cdots,x_n)$，使得 $p(\theta_1 < \theta < \theta_2) = 1-\alpha$，则区间 (θ_1,θ_2) 叫**置信区间**，θ_2 和 θ_1 分别为置信区间的上下限，$1-\alpha$ 叫置信系数，也叫**置信概率**或**置信度**。而 α 是事先给定的一个小正数。它是参数估计不准的概率（假设检验中称之为**检验水平**），一般常给 $\alpha=5\%$ 或 1%。

对区间估计的概念，我们再作一个形象的比喻。我们经常说，上证指数明年大概在 3000 点左右，这句话就可以被看成一个区间估计的问题，如图 1.11 所示。

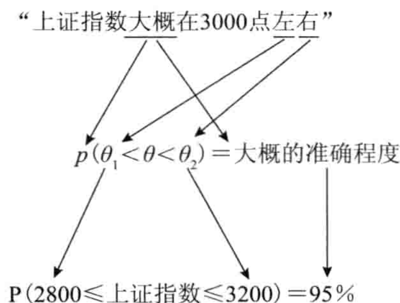


图 1.11

也就是说，明年上证指数有 95% 的可能性在 2800 到 3200 之间。

下面，我们分开不同情况对区间估计的方法给予介绍。

（一）对总体期望值 $E\xi$ 的区间估计

1. 方差已知，对 $E\xi$ 进行区间估计

（1）总体分布未知

对这种情况的处理，需要利用切比雪夫不等式。由于本书已经略去了该不等式的内容，又因为总体分布未知的情况在经济计量分析中并不常见，因而，这里我们直接给出利用切氏不等式推导得出的结论：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



从总体 ξ 中抽取样本 $(x_1,\cdots,x_n)$，令 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，则在 $V\xi$ 已知的情况下，估计所得 $E\xi$ 的置信区间可由下式确定：

$$p(\bar{x}-\sqrt{\frac{V\xi}{\alpha n}}<E\xi<\bar{x}+\sqrt{\frac{V\xi}{\alpha n}})\geq 1-\alpha \tag{1.1}$$

其中，$\bar{x}\pm\sqrt{\frac{V\xi}{\alpha n}}$ 分别是置信区间的上、下限。

这里的 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 是一个随机变量，它随样本的不同而不同，从而对 $\bar{x}$ 的每一个可能值 x，都会有一个置信区间。因此，式 1.1 左边不是随机变量落在某一确定区间内的概率，而是常数 $E\xi$ 被随机区间（$\bar{x}-\sqrt{\frac{V\xi}{\alpha n}},\bar{x}+\sqrt{\frac{V\xi}{\alpha n}}$）盖住的可能性，即平均每 100 次抽样计算得的 100 个置信区间中，至少有（$1-\alpha$）×100 个区间包含 $E\xi$。显然，也会遇到某个区间不包含 $E\xi$ 的情况，但这种情况出现的可能性很小，不超过 α。

进一步还可以看出，置信区间（$\bar{x}-\sqrt{\frac{V\xi}{\alpha n}},\bar{x}+\sqrt{\frac{V\xi}{\alpha n}}$）的长度 $2\sqrt{\frac{V\xi}{\alpha n}}$ 还与样本容量 n 有关。由于总希望这一长度越小越好，因而就要求样本容量 n 越大越好。在实际问题中，如果 n 不可能太大，那么也可以适当降低可靠程度，即适当增大 α，使 $\sqrt{\frac{V\xi}{\alpha n}}$ 变小，以提高区间估计的精度。至于如何在实际问题中选取 n 和 α，要视具体情况而定。

由于总体分布未知，用式 1.1 估计的 $E\xi$ 无法利用 ξ 的分布信息，因而式 1.1 给出的置信区间是比较粗糙的。对某些具体类型的随机变量还可以有更准确的估计形式，下面我们讨论对正态分布总体寻找置信区间的方法。

（2）正态总体

设样本 $(x_1,\cdots,x_n)$ 来自正态总体 $N(\mu,\sigma^2)$，由定理 1.5.13 可知 $U=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，对给定的 α，查标准正态分布表可以确定 μ_α，使 $p(|U|<\mu_\alpha)=1-\alpha$，即

$$p(\left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right|<\mu_a)=1-\alpha$$

$$\text{即 } p(\bar{x}-\frac{\sigma}{\sqrt{n}}\mu_a<\mu<\bar{x}+\frac{\sigma}{\sqrt{n}}\mu_a)=1-a \tag{1.2}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

因此μ的置信度为$1-\alpha$的置信区间是（$\bar{x}-\frac{\sigma}{\sqrt{n}}\mu_\alpha,\bar{x}+\frac{\sigma}{\sqrt{n}}\mu_\alpha$），经查表计算，当$\alpha=0.05$时，$\mu_\alpha=1.96$；当$\alpha=0.01$时，$\mu_\alpha=2.576$。

由于检验水平往往在实际问题中被确定为0.05和0.01，所以$\mu_\alpha=1.96$和$\mu_\alpha=2.576$，这两个μ_α值很常用。对于其它检验水平，请读者自己查表计算μ_α，μ_α为标准正态分布函数表中，函数值$1-\frac{a}{2}$所对应的μ。可见，选取同样的样本，由于这种方法利用了分布的信息，它用切氏不等式所进行的估计要更精确些，即置信区间的长度缩小了。

（3）一般总体大样本下$E\xi$的区间估计

首先介绍数理统计中的中心极限定理，这个定理指出，在很宽的条件下，不是正态总体的一般总体$\xi(E\xi=\mu,D\xi=\sigma^2)$，当样本容量相当大的时候，也有$\bar{x}$渐近地服从正态分布（这个定理的证明本书略）。

根据中心极限定理，大样本情况下，对于一般总体仍用式1.2对$E\xi$进行较精确的区间估计。一般来说，在$n=30$时，就可以把$\bar{x}$看作近似地服从正态分布$N(\mu,\frac{\sigma^2}{n})$，当然$n$再大些更好。

2. 方差未知，对$E\xi$进行区间估计

（1）大样本下，方差未知，对$E\xi$进行区间估计（分布不限）

由于是大样本，根据大数定理和中心极限定理，$V\xi$可用s^2代替，从而仍用式1.2对$E\xi$进行估计。

（2）小样本下，对方差未知的正态总体$E\xi$进行区间估计

设小样本$(x_1,\cdots,x_n)$来自正态总体$N(\mu,\sigma^2)$，由于σ^2未知，用式1.2对$E\xi$（即μ）进行估计已有困难。令

$$T=\frac{\bar{x}-\mu}{s/\sqrt{n}}$$

由于总体$\xi\sim N(\mu,\sigma^2)$，根据定理1.5.15，可知T服从具有$n-1$个自由度的t分布。对于给定的α，查具有$n-1$个自由度的t分布临界值表可确定t_α，使$p(|U|\geq t_\alpha)=\alpha$，即$p\left(\left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right|<t_\alpha\right)=1-\alpha$。

因此，μ的置信区间由下式确定：

$$p\left(\bar{x}-\frac{s}{\sqrt{n}}t_\alpha<\mu<\bar{x}+\frac{s}{\sqrt{n}}t_\alpha\right)=1-\alpha$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

这里，我们实际上是用 s^2 代替了式 1.2 中的 σ^2，从而引起了分布的变化，对此的分析请参见第五节中对定理 1.5.15 的说明。

（二）对总体方差 σ^2 的区间估计

我们只介绍小样本下，正态总体方差 σ^2 的区间估计。

设样本 $(x_1, \cdots, x_n)$ 来自正态总体 $N(\mu, \sigma^2)$，由定理 1.5.14 可知，$Z = \dfrac{(n-1)s^2}{\sigma^2}$ 服从具有 $n-1$ 个自由度的 χ^2 分布。对于给定的 α 查 χ^2 分布临界值表可确定 a 及 b，使

$$p(a < Z < b) = p(a < \frac{(n-1)s^2}{\sigma^2} < b) = 1 - \alpha$$

由此，σ^2 的置信区间由下式确定：

$$p(\frac{(n-1)s^2}{b} < \sigma^2 < \frac{(n-1)s^2}{a}) = 1 - \alpha$$

在确定 a、b 时，一般是取 $p(Z \leq a) = p(Z \geq b) = \dfrac{\alpha}{2}$。

注意：

1. 由以上分析中可以看出，区间估计在方法上就是定理 1.5.13 ~ 1.5.17 的应用。

2. 在进行区间估计时，请切实区分不同情况，应用不同方法。如应分清分布是已知还是未知、是大样本还是小样本等，并应充分利用分布信息。

3. 一般地，α 越大，置信度越低，置信区间长度就越小；反之，就越大。

第八节　通过样本，估计总体（三）——假设检验

一、假设检验的概念

我们称任何一个有关随机变量未知分布的假设为统计假设或简称假设。一个仅涉及到随机变量分布中几个未知参数的假设称为参数假设。如果一个参数假设中的参数完全确定了所要研究的未知分布，那么我们称这种参数假设为单一假设。否则，称之为复合假设。例如，假设 $\mu = 1, \sigma = 2$ 为单一假设，它通过假定 $\mu = 1$ 和 $\sigma = 2$，从而完全确定了一个单

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

一的分布；而假设 $\mu=1$，$\sigma\neq2$ 则为复合假设，这是因为满足这个假设的分布有无数个。

我们这里所说的“假设”只是一个设想，至于它是不是成立，在建立假设时我们并不知道。我们的任务就是通过对样本进行考察，充分利用样本所提供的信息，来确定我们对总体参数的假设是否成立。这样的过程，我们称之为假设检验。

我们称要被检验是否正确的假设为待检假设或零假设（Null Hypothesis），通常用 H_0 表示。检验各种形式的 H_0 是否成立的方法是我们即将讨论的主要内容。

二、两类错误

由于我们作出判断的依据是一组样本，也就是我们在由部分来推测整体。因而假设检验的结果不可能绝对正确，它也可能是错误的。而出现错误可能性的大小，也是以统计规律性为依据的。设 H_0 为待检假设，H_1 是 H_0 的对立假设。在假设检验中，所可能犯的错误通常有两类：

可能犯的第一类错误是：原假设 H_0 符合实际情况，而检验结果把它否定了，这叫**弃真错误**。设 α 是这类错误的概率，那么 $\alpha=p$（否定 $H_0 \mid H_0$ 实际上正确）。

可能犯的第二类错误是：原假设 H_0 不符合实际情况，而检验结果却把它肯定下来了，这叫**取伪错误**。设 β 是这类错误的概率，那么 $\beta=p$（接受 $H_0 \mid H_0$ 不正确）$=p$（接受 $H_0 \mid H_1$ 正确）。

自然，我们希望犯这两类错误的概率越小越好。但对于一定的样本容量 n，一般说来，不能同时做到犯这两类错误的概率都很小。因而我们往往是先固定“犯第一类错误”的概率 α，再考虑如何减小“犯第二类错误”的概率 β。这种 Fix α，Min β 的方法，叫做 Neyman - Pearson 方法。

一般地，我们在进行假设检验前，根据 Neyman - Pearson 方法，总是事先选取一个小正数如 5% 或 1% 作为 α 的确定值，即我们限定当 H_0 事实上正确而我们的检验结果将其否定的概率应该小于 5% 或 1%。根据习惯，我们称 α 为假设检验的显著性水平或检验水平（Significance Level）。

如上所讲，当 α 固定下来后，我们应使 β 尽量地小。β 越小，$1-\beta$ 越大，说明假设检验的结果越可靠。我们称 $1-\beta$ 为假设检验的检验能力（Power of Test）。

由于不可能做到两类错误的概率都很小，因而我们常面临 α 变小则 β 加大和 β 变小则 α 加大这二者之间的抉择。一般来说，在经济计量研究中，我们常愿意使第一类错误的概率尽量小一些，即选择一个较小的 α，从而使当 H_0 正确时，拒绝它的可能性尽量减小。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

三、假设检验与区间估计的关系：置信区间法

首先，我们重新引用当初讨论区间估计的概念时举过的一个例子：我们提到过，“上证指数明年大概在 3000 点左右”这句话可以被看成一个区间估计的问题，如图 1.12 所示。

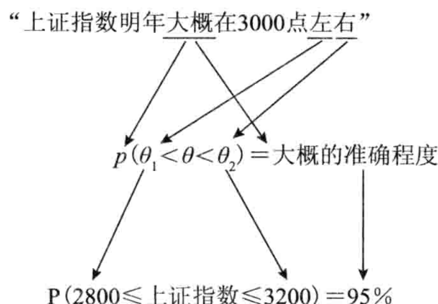


图 1.12

也就是说，上证指数明年有 95% 的可能性会在 2800—3200 点之间。因此，（2800，3200）就是对明年上证指数的一个区间估计，且置信度为 5%。

类似地，我们也举一个关于假设检验的形象例子。比如，某个允许你犯 5% 以下的错误，即以 95% 的正确性来回答如下问题：

上证指数明年在3000点，对吗？
（假设：上证指数明年在3000点；检验：对吗？）

这就是一个假设检验的问题，待检假设 H_0：上证指数明年在 3000 点，而检验水平为 5%。

对比区间估计和假设检验两种情况，我们会发现区间估计实际上给出了一种进行假设检验的方法。比如，当拿到 H_0：上证指数明年在 3000 点（$\alpha=5\%$）后，我们首先对明年的上证指数进行区间估计，得到其在 2800—3200 点之间的概率为 95%。若原假设 H_0 落在置信区间（2800，3200）内，那么我们显然应接受 H_0。这种利用区间估计来进行假设检验的方法称为置信区间法。

对上述例子进一步分析可以发现，给定 5% 的置信度，我们对上证指数进行区间估计，结果为（2800，3200），若原假设落入该区间，我们便接受 H_0，认为明年上证指数应为

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

3000 点，这样我们可能会犯如下一种错误，即明年上证指数实质上为 2600 点，而不是 3000 点，即 H_0 实际上是错误的，而我们却把它接受了，这就是我们上面谈到的第二类错误。而 95% 的置信度，只是保证了当我们应用置信区间法进行假设检验时，在 95% 的情况下，如果 H_0 正确，我们不会拒绝它，即 95% 地防止了假设检验中第一类错误的发生，也就是使检验水平达到了 5%。由此可见，在利用置信区间进行假设检验时，区间估计的置信度 $1-\alpha$ 中的 α，就是假设检验中的检验水平 α。

上述分析也可以使我们进一步弄明白，为什么第一类错误的概率和第二类错误的概率不可能同时变得很小。在置信区间法下，随着检验水平 α 的减小（即第一类错误概率减小），例如从 5%→1%，区间估计的置信度就会加大（从 95%→99%），在区间估计一节中我们已经了解，置信度的加大会使置信区间长度变大，在上例中，比如从（2800，3200）变成（2500，3500），这样就加大了第二类错误发生的概率 β。所以 $\alpha\uparrow \rightarrow \beta\downarrow$，$\alpha$、$\beta$ 不可能同时变小。由以上分析可知，我们以后遇到的各种形式的假设检验，都可以转化为区间估计的问题。

直接用给出的置信区间进行假设检验的方法，虽然比较直观，但不是很方便。下面我们讨论另一种假设检验的方法，这种方法与上例中使用的方法实质相同，但步骤形式略有不同。

在上面，我们实质上是计算 $(\bar{x}-\frac{\sigma}{\sqrt{n}}\mu_\alpha, \bar{x}+\frac{\sigma}{\sqrt{n}}\mu_\alpha)$ 这一 μ 的置信区间，然后看一看 μ_0 是否超出了这个区间。若超出了，则拒绝 H_0，认为 $\mu\neq\mu_0$；否则，接受 H_0。更进一步说，也就是看不等式：

$$\bar{x}-\frac{\sigma}{\sqrt{n}}\mu_\alpha < \mu_0 < \bar{x}+\frac{\sigma}{\sqrt{n}}\mu_\alpha$$

是否成立。若成立，则接受 H_0。

对这个不等式进行变形，有：

$$-\frac{\sigma}{\sqrt{n}}\mu_\alpha < \mu_0-\bar{x} < \frac{\sigma}{\sqrt{n}}\mu_\alpha$$

$$-\mu_\alpha < \frac{\mu_0-\bar{x}}{\sigma/\sqrt{n}} < \mu_\alpha(\mu_\alpha > 0)$$

$$\left|\frac{\mu_0-\bar{x}}{\sigma/\sqrt{n}}\right| < \mu_\alpha$$

可见，不等式 $\bar{x}-\frac{\sigma}{\sqrt{n}}\mu_\alpha < \mu_0 < \bar{x}+\frac{\sigma}{\sqrt{n}}\mu_\alpha$ 等价于不等式 $\left|\frac{\mu_0-\bar{x}}{\sigma/\sqrt{n}}\right| < \mu_\alpha$，二者都可用来

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

判断是否接受 H_0。据此，我们给出已知方差，检验假设 $H_0:\mu=\mu_0$ 这一类问题的另一种检验步骤：

1. 提出待检假设 $H_0:\mu=\mu_0$（μ_0 为已知数）。

2. 先取样本 $x_1,\cdots,x_n$ 的统计量 $U=\dfrac{\bar{x}-\mu_0}{\sigma_0/\sqrt{n}}$（$\sigma_0$ 为已知）：在 H_0 成立条件下，U 应服从标准正态分布。

3. 根据给定的检验水平 α，查表确定临界值 μ_α，使 $p(|U|>\mu_\alpha)=\alpha$

4. 根据样本观察值计算统计量 U 的值并与 μ_α 比较。

5. 下结论：若 $|U|>\mu_\alpha$，则否定 H_0；若 $|U|<\mu_\alpha$，则接受 H_0；若 $|U|=\mu_\alpha$，则不下结论，再进行一次抽样重新检验。

容易看出，这种方法是对直接计算置信区间进行假设检验方法的一种标准化和规范化。其优点是容易记忆，便于使用。这种方式，很像我们日常生活中的其它检验过程，如验血过程，这一点需深入体会。

对这方法，我们可以给出如下理论解释：如果 $H_0:\mu=\mu_0$ 是正确的，那么，我们就可以认为样本 $x_1,\cdots,x_n$ 来自正态总体 $N(\mu_0,\sigma^2)$，根据定理 1.5.13，$\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$，据此选取统计量 $U=\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$，对于给定的 α，可以查表确定 μ_α，使 $p(|U|>\mu_\alpha)=\alpha$，这就是说，如果真有 $U\sim N(0,1)$，即如果 H_0 果真成立，$|U|$ 理论上应小于 μ_α。因此，我们根据样本的观察值计算出 U 的实际值并将其与 μ_α 相比较，就可以确定 H_0 的真假。

由于上述两种方法可以互推，容易看出，二者的实质是相同的，只是分析的角度不同而已。以后，我们将广泛利用第二种方法解决实际中常遇到的各种假设检验问题。

四、假设检验的应用：一个正态总体的假设检验

设总体为 $\xi\sim N(\mu,\sigma^2)$，关于对其参数 μ,σ^2 的假设检验问题，本节介绍以下四种情况：

1. 已知方差 σ^2，检验假设 $H_0:\mu=\mu_0$。
2. 未知方差 σ^2，检验假设 $H_0:\mu=\mu_0$。
3. 未知期望 μ，检验假设 $H_0:\sigma^2=\sigma_0^2$。
4. 未知期望 μ，检验假设 $H_0:\sigma^2\le\sigma_0^2$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

其中 H_0 中的 σ_0^2，μ_0 均为已知数。

前面，我们已经给出了关于第 1 种情况的检验步骤，类似地，这里我们给出其它三种情况的检验步骤并连同第一种情况列表，如表 1－8 所示。这些步骤，读者可用本节三中介绍的方法推导而得。

表 1－8 一个正态总体的假设检验

类型 步骤	已知方差 σ^2 $H_0:\mu=\mu_0$	未知方差 σ^2 $H_0:\mu=\mu_0$	未知期望 μ $H_0:\sigma^2=\sigma_0^2$	未知期望 μ $H_0:\sigma^2\le\sigma_0^2$
1	提出待检假设： $H_0:\mu=\mu_0$ （μ_0 已知）	建立待检假设： $H_0:\mu=\mu_0$ （μ_0 已知）	建立待检假设： $H_0:\sigma^2=\sigma_0^2$	建立待检假设： $H_0:\sigma^2\le\sigma_0^2$
2	选取样本 $x_1,\cdots,x_n$ 的统计量 $U=\dfrac{\bar{x}-\mu_0}{\sigma_0/\sqrt{n}}$（$\sigma_0$ 已知） 若 H_0 成立， $U\sim N(0,1)$	选取样本 $x_1,\cdots,x_n$ 的统计量 $T=\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}$ 若 H_0 成立， $T\sim t(n-1)$	选取样本 $x_1,\cdots,x_n$ 的统计量 $x^2=\dfrac{(n-1)s^2}{\sigma_0^2}$ 若 H_0 成立， $x^2\sim\chi^2(n-1)$	同左
3	根据检验水平 α，查表确定临界值 μ_α，使 $p(\lvert U\rvert>\mu_\alpha)=\alpha$	根据检验水平 α，查表确定临界值 t_α，使 $p(\lvert T\rvert>t_\alpha)=\alpha$	由给定的 α，查表确定 χ_a^2 和 χ_b^2 使： $p(\dfrac{(n-1)s^2}{\sigma^2}>\chi_b^2)=$ $p(\dfrac{(n-1)s^2}{\sigma^2}<\chi_a^2)=\dfrac{\alpha}{2}$	由给定的 α，查表确定 χ_α^2，使： $p(\dfrac{(n-1)s^2}{\sigma_0^2}<\chi_\alpha^2)$ $=\alpha$
4	根据样本观察值计算统计量 U 的实际值 u，与临界值 μ_α 比较	根据样本观察值计算统计量 T 的值并与 t_α 比较	利用样本观察值计算 $\dfrac{(n-1)s^2}{\sigma_0^2}$ 并与 χ_a^2 和 χ_b^2 相比较	利用样本观察值计算 $\dfrac{(n-1)s^2}{\sigma_0^2}$ 并与 χ_α^2 相比较
5	下结论： 若 $\lvert U\rvert>\mu_\alpha$，拒绝 H_0； 若 $\lvert U\rvert\le\mu_\alpha$，接受 H_0； 若 $\lvert U\rvert=\mu_\alpha$ 可再进行一次抽样重新检验	下结论：方法同左。	下结论： 若 $\dfrac{(n-1)s^2}{\sigma^2}>\chi_b^2$ 或 $\dfrac{(n-1)s^2}{\sigma_0^2}<\chi_a^2$，拒绝 H_0； 若 $\dfrac{(n-1)s^2}{\sigma_0^2}$ 在（χ_a^2，χ_b^2）内，接受 H_0	下结论： 若 $\dfrac{(n-1)s^2}{\sigma^2}\ge\chi_\alpha^2$，拒绝 H_0；否则，接受 H_0

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



五、"小概率原理"在假设检验中的应用

数理统计中的"小概率原理"认为：概率很小的事件在一次抽样检验中几乎是不可能发生的。我们进行各种假设检验的依据就是"小概率原理"，如果根据 H_0，推出了一个违背此原理的结论，我们就必须拒绝 H_0 。

事实上，所谓检验水平 α，就是对什么是小概率事件做出的量的规定。如果我们取 $\alpha = 0.05$，那么就意味着，我们认为 0.05 的概率是一个小概率，以此概率发生的事件是一个小概率事件。

最后，我们再强调几点：

1. 数理统计研究的核心问题是如何从样本来推断总体的性质。作为观察者，我们对总体的状况往往是不了解的，我们所能做到的，只是对总体进行随机抽样，从而获得一组样本，再通过对这组样本的研究，进而估计总体的各种属性（参见图 1.2）。样本是总体的一部分，图上的箭头表示了我们的研究方向。我们的各种研究，都是基于样本之上的。

2. 为了描述统计总体，我们引入了随机变量。只有随机变量这样的特殊变量，才能对总体进行全面的描述。

3. 总体就是一个随机变量。

4. 对通过点估法得到的参数估计量，我们常常用假设检验来检验其合理性，同时也用估计量的特性对其进行衡量。一般认为，一个点估计提供的估计量，应当首先具备一个好的估计量所必须具备的特性，如无偏性和有效性；其次应通过对它的显著性检验（即假设检验）；而后才可以被看成总体参数的一个合理估计量。

5. 区间估计和假设检验是从不同角度阐述同一统计方法，具有同质性。

6. 我们在区间估计和假设检验中，广泛地应用了定理 1.5.13 ~ 1.5.17 的结论。这是因为正是这五个定理给出了母体分布和样本分布之间的关系，它们是在我们已知的样本和我们未知的母体之间相互联系的桥梁。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第九节 应用实例：最佳持股周期

最佳持股周期是量化投资中十分重要的一个问题。这里，我们应用数理统计的一些基本方法，对上证指数的最佳持股进行一些较为直接和简单的分析。我们所使用的数据来源于雅虎的金融历史数据库（http：//finance. yahoo. com/q/hp？ s = 000001. SS + Historical + Prices）。样本区间为：1992 年 1 月 2 日至 2013 年 9 月 4 日。我们的目的主要是希望读者能够了解如何在今后的量化投资中进行简单分析，更希望读者能够对此加以改进，对今后的投资有一定的帮助，投资有风险，切勿生搬硬套，囫囵吞枣。

首先，我们需要用 Excel 计算出以下持股周期的回报率，当然为了计算方便我们假设：5 个交易日为一周，20 个交易日为一个月，50 个交易日为一个季度，200 个交易日为一年：

每日回报率　$R1(t) = (P(t) - P(t-1)) / P(t-1)$；

每周回报率　$R5(t) = (P(t) - P(t-5)) / P(t-5)$；

每月回报率　$R20(t) = (P(t) - P(t-20)) / P(t-20)$；

每季度回报率 $R50(t) = (P(t) - P(t-50)) / P(t-50)$；

每年回报率　$R200(t) = (P(t) - P(t-200)) / P(t-200)$。

然后，我们需要计算不同持仓周期回报率的标准差：$\sigma(t)$。在 Excel 中可直接使用 stdev( ) 函数。回报率的均值和标准差的具体计算结果如下：

表 1-9

	R1	R5	R20	R50	R200
均值	0. 000669	0. 003635	0. 015723	0. 042981	0. 216914
标准差	0. 026768	0. 068322	0. 151188	0. 269557	0. 746691

在投资分析中，我们并不是简单将回报率的均值和方差直接进行比较，而是采用夏普比率（Sharpe - ratio），即风险调整后的收益，进行比较：

夏普比率（Sharpe - ratio） = 回报率均值/回报率标准差

表 1-10

	R1	R5	R20	R50	R200
均值	0. 000669	0. 003635	0. 015723	0. 042981	0. 216914
标准差	0. 026768	0. 068322	0. 151188	0. 269557	0. 746691
夏普比率	0. 024976	0. 053199	0. 103994	0. 159453	0. 290501

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



由于存在不同的投资周期，所以，我们要进行比较就必须把不同周期的回报调整为相同时段，这里，我们调整到200天（年）。也就是说，将所有的回报率都转化成年度回报率，即：每日回报率均值×200，每周回报率均值×40，每月回报率均值×10，每季度回报率均值×4。调整后的具体计算结果如表1－11：

表1－11

	R1	R5	R20	R50	R200
均值	0.133711	0.145386	0.157227	0.171926	0.216914
标准差	0.378551	0.432107	0.478099	0.539113	0.746691
夏普比率	0.353217	0.336457	0.328858	0.318905	0.290501

虽然，上面的结果可以进行对比分析，但仍需要考虑交易手续费问题。这里，我们假设每次交易的手续费为千分之一（0.1%），那么每日持股周期的年度手续费为：0.1% * 200 = 20%，每周持股周期的年度手续费为：0.1% * 40 = 4%，每季度持股周期的年度手续费为：0.1% * 4 = 0.4%。具体的计算结果如表1－12：

表1－12

	R1	R5	R20	R50	R200
均值	－0.06629	0.105386	0.147227	0.169926	0.216914
标准差	0.378551	0.432107	0.478099	0.539113	0.746691
夏普比率	－0.17511	0.243888	0.307942	0.315195	0.290501

尽管上面的结果考虑了手续费，但货币是有时间价值的，我们仍需要考虑利率成本。这里，我们假设年利率为4%。考虑利率成本后的计算结果为：

表1－13

	R1	R5	R20	R50	R200
均值	－0.10629	0.065386	0.107227	0.129926	0.176914
标准差	0.378551	0.432107	0.478099	0.539113	0.746691
夏普比率	－0.28078	0.151318	0.224277	0.240999	0.236931

从上面简单的统计分析，我们可以看出，月和季度（甚至年）持股周期的夏普比率较为接近，这就需要进一步对它们的分布进行比较。作为练习，读者可以进一步算出不同持股周期回报率的分布进行比较。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第二章 数据与模型

我们已经知道，在研究量化投资问题时，数字往往是证明观点的证据。基于事实的数据，实际上是我们透过大千世界的外表而探求其运行本质的基础。数据是基础，建立在它之上的是变量，变量之间的相互关系将构成模型。

第一节 数据、模型和变量

一、数据

数据有三种类型：一类是截面数据（Cross - Section Data）；一类是时间序列数据（Time - Series Data）；一类是面板数据（Panel Data），即前面两种数据的组合。

截面数据研究的是某个时间点上的变化情况。例如，表 2 - 1 数据是 2013 年 9 月 6 日上海和深圳证券交易所收盘后股票交易情况的截面数据：

表 2 - 1

	指数	变化	回报率
上证指数	2139.99	+17.56	0.83%
深证成指	8280.33	-5.73	-0.07%
中小板综	5915.46	+26.22	0.45%
沪深 300	2357.78	+16.04	0.69%
沪市 B 股	243.05	+1.59	0.66%

时间序列数据是研究在一定时间范围内的变化情况以及数据之间可能存在的某种联系。下面的数据是 2013 年 8 月 26 日后连续几日上证指数的收盘价：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



表 2-2

时间	上证指数
2013.8.26	2096.47
2013.8.27	2103.57
2013.8.28	2101.30
…	…

这是典型的时间序列数据，每个数据都具有时间性。

面板数据就是时间序列数据和截面数据的合成体，它是既研究某段时间内又研究每个时点上的数据。

在量化投资的研究中，时间序列数据占有很大比重，但是这类数据有很强的周期特点。由于这种周期性因素，对时间序列数据的处理很难直接使用计量经济学所常用的最小二乘法直接回归出直线。如果用最小二乘法回归，会发现真正变动趋势与所得直线之间的差异呈周期性变化，这就是所谓自相关问题。

二、模型

有了数据，当然就要研究其经济关系了。模型是经济关系最直观的表达。模型共有两类：

第一类是单个方程的模型。比如我们在宏观经济研究中，要研究 C_t 与 GDP（Y_t）之间的相互关系，这就是一个时间序列方程，我们可以建立如下方程：

$$C_t = \alpha + \beta Y_t + \xi_t \quad (2.1)$$

第二类是多个方程的模型。例如，收入与消费的关系是相互影响的关系。消费是收入的重要组成部分，对收入产生影响；同时一国收入的多少也会在一定程度上影响着消费。这样，我们可以采用联立方程表示如下：

$$Y_t = C_t + I_t \quad (2.2)$$

其中：C_t 为消费，Y_t 为收入，I_t 为投资。

怎样得到 α、β 呢？能否直接采用最小二乘法对 α、β 进行回归呢？当然不能。我们无法只用第一式进行最小二乘回归而忽略影响第一式中变量变化的第二式。怎么办？现在可以采用间接最小二乘法对 $C_t = \gamma + \delta * I_t$ 进行回归，有 $\hat{\gamma} = \frac{\alpha}{1-\beta}$，$\hat{\delta} = \frac{\beta}{1-\beta}$，据此解得 α，β 的估计值 $\hat{\alpha}$，$\hat{\beta}$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

三、变量

变量是建立模型的基础，它主要分为外生变量（Exogenous Variable）和内生变量（Endogenous Variable）两类。

内生变量是指由经济系统本身决定的变量。系统运行，变量也就随之变化。外生变量是指由经济系统本身无法决定而由外部因素决定的变量。在方程组 $\begin{cases} Y_t = C_t + I_t \\ C_t = \beta * Y_t + \xi_t \end{cases}$ 中，C_t、Y_t 为内生变量，而 I_t 为外生变量。

滞后变量（Lagged Variable），也称前定变量（Predetermined Variable），也属外生变量。例如时间为 t 时，变量 y_t 的前定变量 y_{t-1} 不能改变，因此，y_{t-1} 就是个外生变量。

再如，在国际商品市场上，某种商品的需求函数为：

$$Q_t^D = \alpha - \beta P_t + \varepsilon_t$$

供给函数为：

$$Q_t^s = r + \delta P_{t-1} + V_t$$

供求均衡时：

$$Q_t^s = Q_t^D$$

在上述模型中，P_t 是什么变量呢？如果 P_t 为外生变量，是给定的，那么 Q_t^D 就给定了。但由于 P_{t-1} 也是给定的，从而 Q_t 给定。这样均衡条件 $Q_t^s = Q_t^D$ 很难保证成立。因此，P_{t-1} 影响 Q_t^s 再通过 Q_t^s 影响 P_t，可见 P_t 是内生变量，是根据一系列的相互作用决定的。

第二节　方程形式

进行最小二乘法回归时，我们当然希望模型是线性方程。如果不是，能不能把它变成线性呢？这是本节要论证的问题。

一般而言，主要的方程类型有：

1. 线性方程：$y_t = \alpha + \beta x_t$

2. 对数线性方程：$\ln(y_t) = \alpha + \beta \ln(x_t)$

在已知 x，y 数据后，可以对 x，y 进行数据转换，求得 $\ln(y_t)$，$\ln(x_t)$，然后再把 $\ln(y_t)$，$\ln(x_t)$ 当作变量，进行最小二乘法回归就行了。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



3. 准对数方程：(1) $\ln(y)=\alpha+\beta x$

(2) $y=\alpha+\beta\ln(x)$

已知 x 、y，自然也知道 $\ln(y)$ 、$\ln(x)$，因此分别回归即可。

4. 逻辑方程 (Logistic Model)：

如何处理诸如：$y=\frac{c}{1+e^{\alpha-\beta x}}$ (c 为已知常数) 的方程，将之线性回归？对该方程的处理，需要作以下变形：

$$\frac{y}{c}=\frac{1}{1+e^{\alpha-\beta x}}$$

$$\Rightarrow \frac{c}{y}=1+e^{\alpha-\beta x}$$

$$\Rightarrow \frac{c}{y}-1=e^{\alpha-\beta x}$$

$$\Rightarrow \ln(\frac{c}{y}-1)=\alpha-\beta x$$

现在，把 $\ln(\frac{y}{c}-1)$ 当作新变量，这个变量经过数据转换后是可求得的，这样就可以直接进行线性回归了。

再如，对 $y=\frac{x}{a-bx}$ 的处理如下：

$$y=\frac{x}{a-bx}$$

$$\Rightarrow \frac{1}{y}=\frac{a}{x}-b$$

$$\Rightarrow \frac{1}{y}=a(\frac{1}{x})-b$$

这就可以直接进行线性回归了。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第三节　相关系数

协方差将帮助我们阐述两个随机变量相互关联的程度，参阅图 2.1。

$x-y$ 分布图中，根据 $\bar{x}$、$\bar{y}$ 一般水平可以分割出四个象限，在Ⅰ、Ⅲ象限中 $(x_i-\bar{x})(y_i-\bar{y})$ 是大于 0 的，而在Ⅱ、Ⅳ象限中 $(x_i-\bar{x})(y_i-\bar{y})$ 是小于 0 的。

如果想确定两个变量的关系，由于两个变量都在变化，自然要找到一个标准来衡量。这样，我们就不妨都把其一般水平（$\bar{x}, \bar{y}$）作为标准。

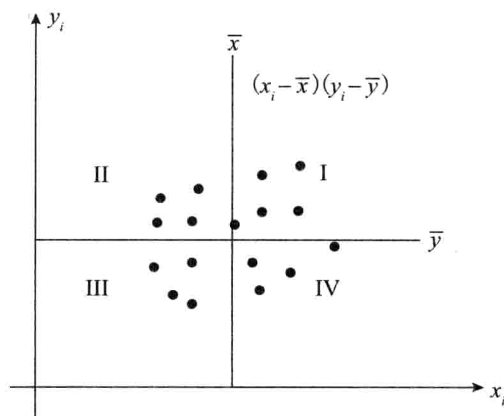


图 2.1

如果大部分点都落在Ⅰ、Ⅲ象限，可能个别点落在Ⅱ、Ⅳ象限中，即落在Ⅰ、Ⅲ内的概率远大于落在Ⅱ、Ⅳ内的概率，说明这两个变量呈正相关变化。协方差的定义就来源于此逻辑。

变量 x、y 的协方差的定义为：

$$\mathrm{cov}(x, y) = E[(x_i-\bar{x})(y_i-\bar{y})] = \sum p_{ij}(x_i-\bar{x})(y_i-\bar{y})$$

一般来说，人们认为 $p_{ij}=\dfrac{1}{n}$，那么，

$$\mathrm{cov}(x, y) = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n}$$

然而协方差有一个缺陷，可以通过下面的例子说明（见表 2－3、表 2－4）。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



表 2－3

y_i	x_i
100	1
320	2
400	3
…	…

表 2－4

y_i	x_i
10000	1
32000	2
40000	3
…	…

其中，表 2－3 中 y_i 的单位为元，表 2－4 中 y_i 的单位为分。

可见，实际上，表 2－3 和表 2－4 中的 x_i 和 y_i 为同一指标。可是如果计算协方差的话，表 2－4 可能比表 2－3 大 100 倍。这就是说，协方差 $\text{cov}(x, y)$ 受变量单位的影响。为了弥补这一缺陷，引入一个辅助指标：相关系数 $\rho_{x,y}$。可以定义为：

$$\rho_{x,y} = \frac{\text{cov}(x,y)}{\sqrt{S_{xx}}\ \sqrt{S_{yy}}}$$

其中：$S_{xx} = \frac{1}{n}\sum(x_i - \bar{x})^2 = \frac{1}{n}(\sum x_i^2 - n\bar{x}^2)$，

$$S_{yy} = \frac{1}{n}\sum(y_i - \bar{y})^2 = \frac{1}{n}(\sum y_i^2 - n\bar{y}^2)。$$

这样，ρ 就不受单位影响了。

如果 x，y 存在线性关系：$y_i = a + bx_i$，$\rho_{x,y}$ 会怎样呢？

由于

$$y_i = a + bx_i$$

$$\Rightarrow \sum y_i = na + b\sum x_i$$

$$\Rightarrow \frac{\sum y_i}{n} = a + b\frac{\sum x_i}{n}$$

$$\Rightarrow \bar{y} = a + b\bar{x}$$

用 $y_i = a + bx_i$ 减去 $\bar{y} = a + b\bar{x}$ 有 $y_i - \bar{y} = b(x_i - \bar{x})$

那么，

$$\rho_{x,y} = \frac{\text{cov}(x,y)}{\sqrt{S_{xx}}\ \sqrt{S_{yy}}} = \frac{\sum p_{ij}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}\ \sqrt{\frac{1}{n}\sum(y_i - \bar{y})^2}}$$

由于 p_{ij} 通常为 $\frac{1}{n}$，则有：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



$$\rho_{x,y}=\frac{\frac{1}{n}b\sum(x_i-\bar{x})^2}{\frac{1}{n}|b|\sum(x_i-\bar{x})^2}=\pm 1$$

也就是说，如果 x，y 存在线性关系，则 $|\rho_{x,y}|=1$，只要 x，y 在一条直线上，无论 x，y 值如何变，相关系数总是 1 或 -1，如图 2.2 所示：

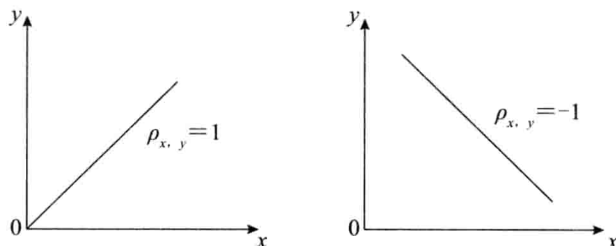


图 2.2

如果 x，y 不是线性，则有 $-1<\rho_{x,y}<1$。

$\rho_{x,y}$ 能否为 0 呢？图 2.3 说明了 $\rho_{x,y}=0$ 的情况。

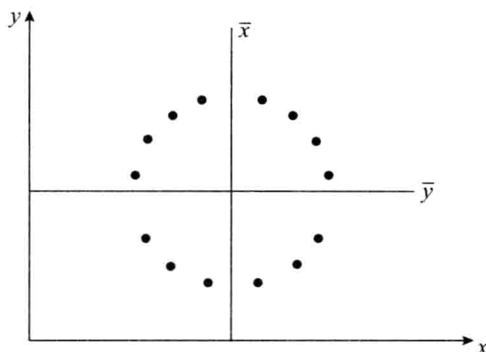


图 2.3

这样，我们可以得到结论：$|\rho_{x,y}|$ 越接近于 1，x，y 之间的相互关联度就越强。

当然，相关系数也有不足之处，它只能说明 x，y 之间是否存在线性关系；但是它无法表明是 x 影响 y，还是 y 影响 x，或是 x，y 相互影响，或是另外因素引起了 x，y 共同变化。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第三章

最小二乘法：一元线性回归

前面谈到的时间序列数据和截面数据，在下面的讨论中将会显示它的重要性。假如我们对两个变量 x 和 y 的关系十分感兴趣，为了精确地描述这种关系，我们需要利用每个变量的观察值来建立起这种关系的详细数学形式。我们首先考察的是 x 和 y 的关系是否为线性，若是线性关系，则这种关系将由一条直线描述。在肯定 x 和 y 的关系是线性关系的前提下，我们的目的是找到一条能够最好地描述 x，y 之间的这种线性关系的直线。

比如，我们想研究股价变动与交易量之间的关系，首先，我们可以得到以下 8 个观察值，如图 3.1 所示。

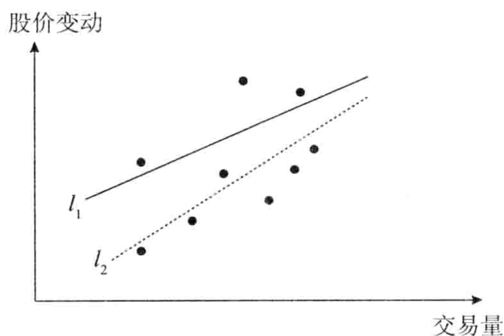


图 3.1

从图 3.1 中可以看出，我们能找到许多拟合直线，如 l_1 和 l_2，但最好的方法是找一条能使这些点到直线上相应各点的垂直坐标距离之和最小的直线。

在研究点与直线的距离问题上会遇到三种情况：点到直线的水平距离，点到直线上相应各点的垂直坐标距离和点到直线的距离，如图 3.2 所示：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

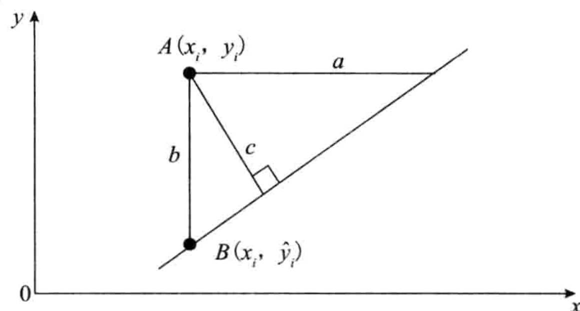


图 3.2

图中：a 为点到直线的水平距离；b 为点到直线上点的垂直坐标距离；c 为点到直线的距离。

本章中涉及的最小二乘法中的距离，指的是点到直线上点的垂直坐标距离，最小二乘法拟合直线是使得点到直线上点的垂直坐标距离最小的那条直线。在此基础上建立起来的最小二乘法，将在本书中大量运用。但这并不排除对其它估计法的研究，如最大似然法等。

最小二乘法无非是基于这样的思路：在许多点中试图能找到一条直线来描述这些点的趋势或特性。这条直线的求法是使各点到该拟合直线上相应各点的垂直坐标距离平方和达到最小。最小二乘法目的是找到一条能最好地描述 y_i 与 x_i 的线性关系的直线方程。如图3.2所示，图中 $A(x_i,y_i)$ 为原始点，$B(x_i,\hat{y}_i)$ 为拟合直线上与之对应的点。二者之间的垂直坐标距离为 $\hat{\varepsilon}_i = y_i - \hat{y}_i$，也即为 A 点到拟合直线的垂直坐标距离。将所有这些点的垂直坐标距离平方后相加，运用求极值的方法使之最小：$\text{Min}\sum\hat{\varepsilon}_i^2$，而 $\text{Min}\sum\hat{\varepsilon}_i^2 = \sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$，其中 y_i 为真实点纵坐标。因此问题转化成求 $\sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$ 的最小值。在已给的样本中，y_i 和 x_i 由于已知，就可以采用求极值的方法求出 α、β 的拟合值：

$$\frac{\partial\ (\text{Min}\sum\hat{\varepsilon}_i^2)}{\partial\ \hat{\alpha}} = -2\sum(y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \tag{3.1}$$

$$\frac{\partial\ (\text{Min}\sum\hat{\varepsilon}_i^2)}{\partial\ \hat{\beta}} = -2\sum(y_i - \hat{\alpha} - \hat{\beta}x_i)x_i = 0 \tag{3.2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{3.3}$$

$$\hat{\beta} = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sum x_i^2 - n\,(\bar{x})^2} \tag{3.4}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

同时，我们还可以从中得到如下结论：

从式 3.3 可以得到：$\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}$ (3.5)

由于 $y_i - \hat{\alpha} - \hat{\beta}x_i = y_i - \hat{y}_i = \hat{\varepsilon}_i$，从式 3.2 中可知 $\sum(y_i - \hat{\alpha} - \hat{\beta}x_i)x_i = 0$，这样可以得到

$$\sum \hat{\varepsilon}_i x_i = 0 \tag{3.6}$$

式 3.5、式 3.6 两点结论有助于我们进一步讨论拟合直线的有关性质。

第一节 拟合直线性质

到目前为止，这条已经求出来的拟合直线仍然是一条相对“模糊”的直线，我们对其所知甚少。如果能得到它的一些性质，将有助于我们更进一步地了解它的若干特征及其作用范围。这条拟合直线具有以下性质：

一、估计残差和为零

即：$\sum \hat{\varepsilon}_i = 0$。

由最小二乘法已经保证了 $\sum \hat{\varepsilon}_i^2$ 最小，事实上，我们就是根据它得出拟合直线形状。但 $\sum \hat{\varepsilon}_i^2$ 最小并不能保证 $\sum \hat{\varepsilon}_i = 0$，这是两回事。因此，要得到 $\sum \hat{\varepsilon}_i = 0$，还需要进一步证实。

证明：第一步（我们欲证 $\sum \hat{\varepsilon}_i = 0$，首先就要创造出 $\sum \hat{\varepsilon}_i$ 来），由于 $y_i = \hat{\alpha} + \hat{\beta}x_i + \hat{\varepsilon}_i$（这个 y_i 是样本真实值），进行累加后可得到 $\sum y_i = n\hat{\alpha} + \hat{\beta}\sum x_i + \sum \hat{\varepsilon}_i$。

第二步，上式两边同时除以 n，可得 $\frac{\sum y_i}{n} = \hat{\alpha} + \frac{\hat{\beta}\sum x_i}{n} + \frac{\sum \hat{\varepsilon}_i}{n}$。

由于 $\frac{\sum y_i}{n} = \bar{y}$，$\frac{\sum x_i}{n} = \bar{x}$，等式就化为：$\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x} + \frac{\sum \hat{\varepsilon}_i}{n}$。

想一想我们从最小二乘法过程中得到的式 3.5，$\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}$ 很显然必定有 $\frac{\sum \hat{\varepsilon}_i}{n} = 0$，所以也就有 $\sum \hat{\varepsilon}_i = 0$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



二、y_i 的真实值和拟合值 $\hat{y}_i$ 有共同的均值

即：y_i 的均值 = $\hat{y}_i$ 的均值，也即 $\bar{y} = \bar{\hat{y}}$。

证明：同证明性质一的方法一样，我们也必须首先创造出二者的均值来。由 $y_i = \hat{y}_i + \hat{\varepsilon}_i$ 可以得到累加后的等式：$\sum y_i = \sum \hat{y}_i + \sum \hat{\varepsilon}_i$。

由性质一已知 $\sum \hat{\varepsilon}_i = 0$，那么 $\sum y_i = \sum \hat{y}_i$ 是成立的，两边同时除以 n 来求均值，就会得到 $\bar{y} = \bar{\hat{y}}$。

三、估计残差与自变量不相关

即：$\text{cov}(x,\hat{\varepsilon}) = 0$ 或 $\rho_{x,\hat{\varepsilon}} = 0$。

证明：根据定义，我们知道：$\text{cov}(x,\hat{\varepsilon}) = \frac{1}{n}\sum(x_i - \bar{x})(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})$。

第一步，我们来看一看 $\bar{\hat{\varepsilon}}$。

$\bar{\hat{\varepsilon}}$ 等于 $\frac{\sum \hat{\varepsilon}_i}{n}$，而 $\sum \hat{\varepsilon}_i$ 已经证明了是等于零的，所以 $\bar{\hat{\varepsilon}} = \frac{\sum \hat{\varepsilon}_i}{n} = 0$。

那么：$\frac{1}{n}\sum(x_i - \bar{x})(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}) = \frac{1}{n}\sum(x_i - \bar{x})\hat{\varepsilon}_i$。

第二步，把上式右边进行展开后可以得到 $\frac{1}{n}\sum(x_i - \bar{x})\hat{\varepsilon}_i = \frac{1}{n}(\sum x_i\hat{\varepsilon}_i - \bar{x}\sum \hat{\varepsilon}_i)$。

现在再考虑我们已经得到的式 3.6，$\sum x_i\hat{\varepsilon}_i = 0$。

同时拟合直线的性质一又告诉我们 $\sum \hat{\varepsilon}_i = 0$，这样整个式子就变成

$$\frac{1}{n}\sum(x_i\hat{\varepsilon}_i - \bar{x}\hat{\varepsilon}_i) = \frac{1}{n}[\sum x_i\hat{\varepsilon}_i - \bar{x}\sum \hat{\varepsilon}_i] = \frac{1}{n}(0 - \bar{x}\cdot 0) = 0$$

也就是说，$\text{cov}(x,\hat{\varepsilon}) = 0$；同时，$\rho_{x,\hat{\varepsilon}} = \frac{\text{cov}(x,\hat{\varepsilon})}{\sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2 \cdot \frac{1}{n}\sum(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})^2}} = 0$

四、估计残差与拟合值不相关

即：$\text{cov}(\hat{y},\hat{\varepsilon}) = 0$ 或 $\rho_{x,\hat{\varepsilon}} = 0$。

证明：根据定义，我们可知：$\text{cov}(\hat{y},\hat{\varepsilon}) = \frac{1}{n}\sum(\hat{y}_i - \bar{\hat{y}})(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



首先，我们看看 $\bar{\hat{y}}$ 和 $\bar{\hat{\varepsilon}}$。我们在性质二里证明了 $\hat{y}_i$ 与 y_i 的均值是相等的，即：$\bar{y}=\bar{\hat{y}}$。从性质一我们知道 $\bar{\hat{\varepsilon}}=0$，于是上式就变成了 $\mathrm{cov}(\hat{y},\hat{\varepsilon})=\frac{1}{n}\sum(\hat{y}_i-\bar{y})\hat{\varepsilon}_i$

现在将它展开：

$$\begin{aligned}\mathrm{cov}(\hat{y},\hat{\varepsilon})&=\frac{1}{n}\sum(\hat{y}_i-\bar{y})\cdot\hat{\varepsilon}_i=\frac{1}{n}(\sum\hat{y}_i\hat{\varepsilon}_i-\bar{y}\sum\hat{\varepsilon}_i)\\&=\frac{1}{n}[\sum(\hat{\alpha}+\hat{\beta}x_i)\hat{\varepsilon}_i-\bar{y}\times 0]\ (\text{由于}\ \hat{y}_i=\hat{\alpha}+\hat{\beta}x_i,\ \sum\hat{\varepsilon}_i=0)\\&=\frac{1}{n}\sum(\hat{\alpha}+\hat{\beta}x_i)\hat{\varepsilon}_i\\&=\frac{1}{n}(\sum\hat{\alpha}\hat{\varepsilon}_i+\sum\hat{\beta}x_i\hat{\varepsilon}_i)\\&=\frac{1}{n}(\hat{\alpha}\sum\hat{\varepsilon}_i+\hat{\beta}\sum x_i\hat{\varepsilon}_i)\end{aligned}$$

我们同证明性质三时一样，又一次用到了 $\sum\hat{\varepsilon}_i=0$，$\sum x_i\hat{\varepsilon}_i=0$ 从这里可以看出它们给了我们很多帮助，几乎无处不在。利用这两个结果，整个式子最终变成：

$$\mathrm{cov}(\hat{y},\hat{\varepsilon})=\frac{1}{n}(\hat{\alpha}\times 0+\hat{\beta}\times 0)=0$$

从 $\mathrm{cov}(\hat{y},\hat{\varepsilon})=0$，我们很容易根据定义得到：$\rho_{\hat{y},\hat{\varepsilon}}=\dfrac{\mathrm{cov}(\hat{y},\hat{\varepsilon})}{\sqrt{\frac{1}{n}\sum(\hat{y}_i-\bar{\hat{y}}_i)^2\cdot\frac{1}{n}\sum(\hat{\varepsilon}_i-\bar{\hat{\varepsilon}})^2}}=0$。

以上我们着重讨论了拟合直线的若干性质。现在我们来看看这些性质究竟说明了什么？

性质一说明了 $\hat{\varepsilon}_i$ 的特点：$\sum\hat{\varepsilon}_i=0$。

性质二说明了 y_i 和 $\hat{y}_i$ 的关系：它们有同样的均值：$\bar{y}=\bar{\hat{y}}$。

性质三说明了 x_i 和 $\hat{\varepsilon}_i$ 的关系：它们不相关，$\mathrm{cov}(x,\hat{\varepsilon})=0$。

性质四说明了 $\hat{y}_i$ 和 $\hat{\varepsilon}_i$ 的关系：它们不相关，$\mathrm{cov}(\hat{y},\hat{\varepsilon})=0$。

最后应指出的是，无论 y_i 和 x_i 的关系如何，只要是采用最小二乘法得到拟合直线，都具有上面的这些性质。同时，我们还应该注意到，这些性质中有许多是关于拟合直线残差 $\hat{\varepsilon}_i$ 的，如 $\sum\hat{\varepsilon}_i=0$，$\mathrm{cov}(x,\hat{\varepsilon})=0$，$\mathrm{cov}(\hat{y},\hat{\varepsilon})=0$。但这与后面古典模型有关 ε_i 的假设是有区别的，ε_i 是母体残差，而 $\hat{\varepsilon}_i$ 是样本的拟合残差。注意到这种区别，在后面的学习中颇为重要。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第二节　拟合优度的评价

我们从许多杂乱无章的点中拟合出了一条有规律的直线，并且进一步讨论了这条直线的一些性质。这样使我们对这条直线的认识变得更加清晰。但是这条直线究竟能够对这些点之间的关系反映到何种程度，是否能反映这些点相互之间的关系或趋势。换句话说，拟合的好坏究竟怎样，还需要我们去进一步深入研究。

我们经常使用三个指标来帮助测定拟合的好坏。这三个指标是：回归平方和（*RSS*）、残差平方和（*ESS*）、离差平方和（*TSS*）。

一、有关 *RSS*、*TSS*、*ESS* 的定义

定义 3.2.1　$RSS = \sum (\hat{y}_i - \bar{y})^2$

它表示了 y 拟合值的离散程度，即与 y 均值的离散程度如何。

定义 3.2.2　$TSS = \sum (y_i - \bar{y})^2$

它表示真实值 y_i 本身的离散程度，即 y_i 值围绕其样本均值的残值平方和。注意 $TSS/(n-1)$ 是 y_i 的样本方差。

定义 3.2.3　$ESS = \sum (y_i - \hat{y}_i)^2 = \sum \hat{\varepsilon}_i^2$

表示真实值与拟合值之间的差距，实际为 y_i 离 $\hat{y}_i$ 的距离。

如图 3.3 所示，该图较为直观地展示出 *TSS*、*RSS*、*ESS* 对应的方位。但是注意，图中仅表示了一个 y_i 值，而 *RSS*、*TSS*、*ESS* 是所有点的这些距离的平方和。

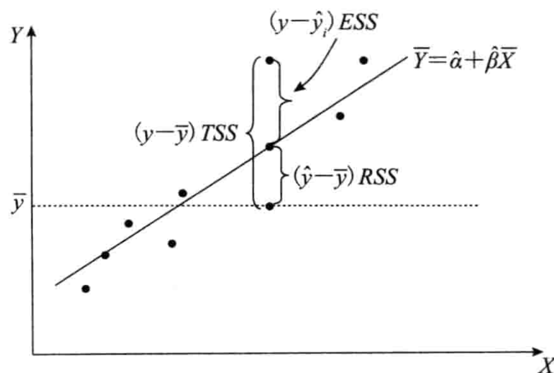


图 3.3

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

二、RSS、ESS 和 TSS 的相互关系

从图 3.3 中可以看出似乎 $TSS = ESS + RSS$，但是应该注意，根据定义可知，TSS、ESS、RSS 均为所有点的这些距离的平方和。因此它们之间的这种关系还要进一步证明，就好像我们知道 $c = a + b$，却不能肯定 $c^2 = a^2 + b^2$ 一定成立一样。

由定义可知，$TSS = \sum (y_i - \bar{y})^2$，对之加以变形后展开：

$$\begin{aligned}TSS &= \sum (y_i - \bar{y})^2 \\ &= \sum [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\ &= \sum [(y_i - \hat{y}_i)^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + (\hat{y}_i - \bar{y})^2] \\ &= \sum (y_i - \hat{y}_i)^2 + 2\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum (\hat{y}_i - \bar{y})^2\end{aligned}$$

已知 $ESS = \sum (y_i - \hat{y}_i)^2$，$RSS = \sum (\hat{y}_i - \bar{y})^2$，因此 $TSS = ESS + RSS + 2\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$，如果 $2\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$ 能等于零，就是最好不过了。

下面，我们来证明 $\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$ 。

$$\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum \hat{\varepsilon}_i(\hat{y}_i - \bar{y}) = \sum \hat{\varepsilon}_i\hat{y}_i - \bar{y}\sum \hat{\varepsilon}_i$$

一个绝妙的想法是解题的关键，$(y_i - \hat{y}_i) = \hat{\varepsilon}_i$ 是基础。

现在我们又可以像证明拟合直线性质四那样如法炮制了。

将 $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$，$\sum \hat{\varepsilon}_i = 0$ 代入上式中：

$$\begin{aligned}&\sum \hat{\varepsilon}_i\hat{y}_i - \bar{y}\sum \hat{\varepsilon}_i \\ &= \sum \hat{\varepsilon}_i(\hat{\alpha} + \hat{\beta}x_i) - \bar{y} \times 0 \\ &= \sum \hat{\alpha}\hat{\varepsilon}_i + \sum \hat{\varepsilon}_i\hat{\beta}x_i \\ &= \hat{\alpha}\sum \hat{\varepsilon}_i + \hat{\beta}\sum \hat{\varepsilon}_i x_i \\ &= \hat{\alpha} \times 0 + \hat{\beta} \times 0 \\ &= 0\end{aligned}$$

这里我们再一次用到了拟合直线的性质：$\sum \hat{\varepsilon}_i = 0$ 以及我们在前面得到的式 3.6 的结果。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



这样，我们就得到 $\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$ 的结果。于是，我们可以肯定地说，RSS、ESS 和 TSS 之间的关系为：$TSS = ESS + RSS$。

现在，我们根据这一等式引入一个新的测定拟合优度的指标：相关系数 R^2。

$TSS = ESS + RSS \Rightarrow \dfrac{ESS}{TSS} + \dfrac{RSS}{TSS} = 1$，那么，$\dfrac{RSS}{TSS} = 1 - \dfrac{ESS}{TSS}$。

我们定义：$R^2 = \dfrac{RSS}{TSS} = 1 - \dfrac{ESS}{TSS}$。

从定义可以看到，TSS 表示 y_i 到 $\bar{y}$ 的离差平方和，RSS 则表示回归直线能帮助减少真实值与均值的偏差程度。相关系数 R^2 作为一个相对量，是反映回归直线所能减少偏离程度的一个相对指标。所以 R^2 越高越好，即拟合直线对 y_i 随机变动的解释程度越大越好，R^2 的范围在 0—1 之间。

当 $TSS = RSS$ 时，说明所有偏差都能够为回归直线所解释，拟合直线过所有的实际点，这将是个十分完美的事情，如图 3.4 所示。

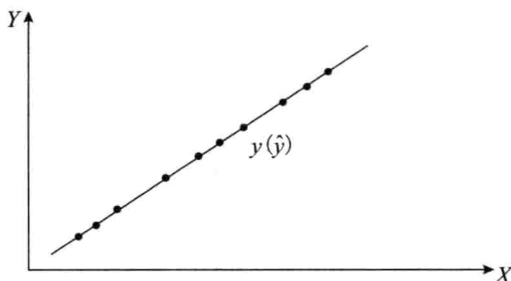


图 3.4

在图 3.4 中，$TSS = RSS$，即 $R^2 = 1$，所有实际点都落到了拟合直线上，滴“点”不漏。

在实际中，这种情况是不可能发生的。拟合直线只能解释 y_i 到 $\bar{y}$ 离差平方和中的一部分。由于相关系数 R^2 表示拟合直线能解释的部分占全部离差的比例，所以我们说 R^2 越趋近于 1 越好。

三、R^2 和 $\rho^2_{y,\hat{y}}$ 的关系

我们先求 $\rho^2_{y,\hat{y}}$，计算公式为：$\rho^2_{y,\hat{y}} = \dfrac{\frac{1}{n}\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\frac{1}{n}\sum (y_i - \bar{y})^2 \frac{1}{n}\sum (\hat{y}_i - \bar{y})^2}$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



毫无疑问，分母 $\frac{1}{n}\sum(y_i-\bar{y})^2\frac{1}{n}\sum(\hat{y}_i-\bar{y})^2=\frac{1}{n}TSS\cdot\frac{1}{n}RSS$

我们来看分子。由拟合直线性质二，$\bar{y}=\bar{\hat{y}}$ 可知

$$
\begin{aligned}
&\sum(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})\\
=&\sum(y_i-\bar{y})(\hat{y}_i-\bar{y})\\
=&\sum(\hat{y}_i+\hat{\varepsilon}_i-\bar{y})(\hat{y}_i-\bar{y})\\
=&\sum[(\hat{y}_i-\bar{y})+\hat{\varepsilon}_i](\hat{y}_i-\bar{y})\\
=&\sum(\hat{y}_i-\bar{y})^2+\sum\hat{\varepsilon}_i(\hat{y}_i-\bar{y})\\
=&RSS+\sum\hat{\varepsilon}_i(\hat{y}_i-\bar{y})
\end{aligned}
$$

请看 $\sum\hat{\varepsilon}_i(\hat{y}_i-\bar{y})$ 像什么？这里需要使用一个技巧。我们已经知道 $\bar{\hat{\varepsilon}}=0$ 了，所以为了得到 $\sum\hat{\varepsilon}_i(\hat{y}_i-\bar{y})=0$，可以先利用拟合直线的性质四：$\mathrm{cov}(\hat{\varepsilon},\hat{y})=0$，具体步骤如下：

$$
\begin{aligned}
&\sum\hat{\varepsilon}_i(\hat{y}_i-\bar{y})\\
=&\sum(\hat{\varepsilon}_i-\bar{\hat{\varepsilon}})(\hat{y}_i-\bar{y})\\
=&n\times\frac{1}{n}\sum(\hat{\varepsilon}_i-\bar{\hat{\varepsilon}})(\hat{y}_i-\bar{y})\\
=&n\times\mathrm{cov}(\hat{\varepsilon},\hat{y})=0
\end{aligned}
$$

则：$\sum(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})=RSS+n\times\mathrm{cov}(\hat{\varepsilon},\hat{y})=RSS$

于是分子部分就变为 $[\frac{1}{n}\sum(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})]^2=(\frac{1}{n}RSS)^2$。

因此

$$
\begin{aligned}
\rho^2_{y,\hat{y}}&=\frac{[\frac{1}{n}\sum(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})]^2}{\frac{1}{n}\sum(y_i-\bar{y})^2\frac{1}{n}\sum(\hat{y}_i-\bar{y})^2}\\
&=\frac{(\frac{1}{n}RSS)^2}{(\frac{1}{n}TSS)(\frac{1}{n}RSS)}\\
&=\frac{RSS}{TSS}=R^2
\end{aligned}
$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

可见 $\rho^2_{y,\hat{y}}$ 和 R^2 一样，也表明拟合的 $\hat{y}_i$ 与实际的 y_i 的相关程度。$\rho^2_{y,\hat{y}}$ 越大，说明 $\hat{y}_i$ 拟合得越好。

我们使用最小二乘法来探讨客观现实中一些点的规律，并且希望能够用一条易于观察的富有代表性的直线来描述和表示这些点之间的相互关系。于是，采用最小二乘法，我们可以得到一条拟合直线，并且由此可以得知该拟合直线的一些性质。但我们得到的这条拟合直线究竟能不能说明这些点的特征？说明的精确程度如何？还需我们运用一些指标对这条拟合直线进行衡量（见图 3.5）。

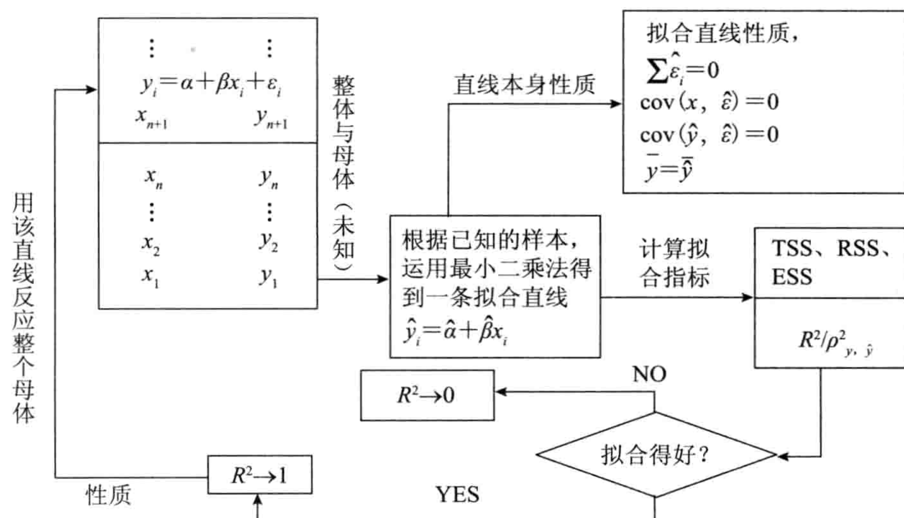


图 3.5

在用最小二乘法进行回归时，用不用常数项会有很大区别：不使用常数项意味着回归出的直线永远经过原点。这极大地限制了直线的拟合作用，不能较好地反映真实情况，远不及带常数项的回归方程灵活。因此，带常数项的要比不带常数项的直线拟合得好。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第三节　一元线性回归中的计算问题

前面两节中，我们主要探讨了一元线性回归中 α、β 的估计量，拟合直线的性质以及拟合优度的评价问题。本节将着重讨论有关 α、β 估计量的具体计算方法。由上面的分析可知，α、β 的估计量为

$$\hat{\beta} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

计算 $\hat{\alpha}$、$\hat{\beta}$ 的最简便易行的方法是采取列表的方法，具体如表 3－1：

表 3－1

①	②	③＝①－$\bar{y}$	④＝②－$\bar{x}$	⑤＝③×④	⑥＝④^2
y_i	x_i	$y_i - \bar{y}$	$x_i - \bar{x}$	$(y_i - \bar{y})(x_i - \bar{x})$	$(x_i - \bar{x})^2$
y_1 y_2 ⋮ y_n	x_1 x_2 ⋮ x_n	$(y_1 - \bar{y})$ $(y_2 - \bar{y})$ ⋮ $(y_n - \bar{y})$	$(x_1 - \bar{x})$ $(x_2 - \bar{x})$ ⋮ $(x_n - \bar{x})$	$(y_1 - \bar{y})(x_1 - \bar{x})$ $(y_2 - \bar{y})(x_2 - \bar{x})$ ⋮ $(y_n - \bar{y})(x_n - \bar{x})$	$(x_1 - \bar{x})^2$ $(x_2 - \bar{x})^2$ ⋮ $(x_n - \bar{x})^2$
$\sum y_i$	$\sum x_i$			$\sum (y_i - \bar{y})(x_i - \bar{x})$	$\sum (x_i - \bar{x})^2$
$\bar{y}$	$\bar{x}$				

⑦＝$\hat{\alpha} + \hat{\beta}x_i$	⑧＝③^2	⑨＝①－⑦	⑩＝⑨^2
$\hat{y}_i$	$(y_i - \bar{y})^2$	$(y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
$\hat{y}_1$ $\hat{y}_2$ ⋮ $\hat{y}_n$	$(y_1 - \bar{y})^2$ $(y_2 - \bar{y})^2$ ⋮ $(y_n - \bar{y})^2$	$(y_1 - \hat{y}_1)$ $(y_2 - \hat{y}_2)$ ⋮ $(y_n - \hat{y}_n)$	$(y_1 - \hat{y}_1)^2$ $(y_2 - \hat{y}_2)^2$ ⋮ $(y_n - \hat{y}_n)^2$
	$\sum (y_i - \bar{y})^2$		$\sum (y_i - \hat{y}_i)^2$

表 3－1 中，第①、②列为原始数据。据此，我们计算出 $\bar{y}$，$\bar{x}$ 及第③和④列。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

由上表中的第⑤列和第⑥列，我们可以计算出 $\hat{\beta} = \dfrac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$。

再由第①②列，可知 $\bar{y}$、$\bar{x}$，因此，可计算出 $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$。

我们根据已求出的 $\hat{\alpha}$、$\hat{\beta}$，可计算出第⑦列的 $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$。

据此，我们还可以进一步计算出拟合优度 R^2，具体步骤如下：

首先，计算出第③列的 $y_i - \bar{y}$，据此计算第⑧列 $(y_i - \bar{y})^2$。这样，我们对第⑧列进行求和，就可以得到：$TSS \equiv \sum (y_i - \bar{y})^2$。

同样，通过第①列和第⑦列，我们可以计算出第⑨列 $\hat{\varepsilon}_i$，同理，我们对第⑨列平方后，就可以得到第⑩列，对第⑩列求和就可以得到：$ESS \equiv \sum (y_i - \hat{y}_i)^2$。

据此，我们可以计算出 $R^2 = 1 - \dfrac{ESS}{TSS}$。

对于较为复杂的数字，我们可以采用 EXCEL 进行相应的计算。我们也可以采用C ++，GAUSS 和 Visual BASIC 等计算机高级语言，编写相应的通用程序，感兴趣的读者可自己进行一些尝试。像 TSP，EViews 等回归分析软件也大多采用类似算法。

第四节　应用案例：对资产定价模型（CAPM）的简单检验

资本资产定价模型（CAPM）是现代金融理论的重要组成部分，对现代金融理论产生了很大的影响。下面将利用资本资产定价模型对浦发银行这支股票进行实证分析。

资本资产定价模型是夏普（William Sharpe）等人在 1965 年提出，被简称为 CAPM 模型。CAPM 模型以简洁的形式和易于操作的优势在诸多方面得到了广泛应用，例如：股票收益预测、证券组合表现评价、事件研究分析、证券股价及确定资本成本等等。

20 世纪 70 年代开始，人们对 CAPM 进行了大量的实证检验。早期的检验结果表明，CAPM 模型在西方成熟的股票市场中是有效的，平均股票收益与整个市场平均收益是呈正线性相关关系的。但从 20 世纪 80 年代后，对 CAPM 模型的实证检验结果发生了变化，多数检验结果并不支持 CAPM 模型了，平均股票收益与整个市场平均收益的这种正相关关系在 20 世纪 70 年代后的数据中就消失了。与此同时，许多其他因素被发现对于股票收益具有显著解释能力，比如考虑了是否存在其他因素能够解释横截面上的差异等。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



国内学者从 20 世纪 90 年代中期开始也用 CAPM 模型在中国的股票市场上进行实证检验和分析。下面将利用简单回归分析方法，对 2012 年 12 月至 2013 年 9 月浦发银行的月度数据进行资本资产定价模型的初步实证分析研究。

CAPM 模型假定投资者能够以无风险收益率借贷，其 CAPM 形式为：

$$E(R_i) - R_f = \alpha + \beta[E(R_m) - R_f]$$

其中：$E(R_i)$ 为第 i 项资产的期望收益率；$E(R_m)$ 为有效市场组合的期望收益率；R_f 为无风险资产的收益率。

下面将对上证综合指数的月收益率与浦发银行的月收益率作简单回归，来估计这只股票的 β 系数，具体数据来源于雅虎网站，数据的具体定义如下：

$E(R_i)_t$ 表示浦发银行在 t 时间的月度收益率；

$E(R_m)_t$ 表示上证综指在 t 时间的月度收益率；

R_f 是 t 时间的无风险收益率，这里简单假设为 0.2%；很多研究 CAPM 模型的文章中，无风险收益都近似利用了三个月居民定期存款利率，但是为了简便起见，这里假设无风险月收益率为 0.2%；

α，β 为估计的系数。

其具体数据见表 3－2：

表 3－2

	上证综指	浦发银行
2012－11－1	1980.12	6.9
2012－12－3	2269.13	9.36
2013－1－4	2385.42	10.92
2013－2－1	2365.59	10.51
2013－3－1	2236.62	9.57
2013－4－1	2177.91	9.32
2013－5－2	2300.59	9.92
2013－6－3	1979.21	8.28
2013－7－1	1993.8	7.86
2013－8－1	2098.38	8.99
2013－9－2	2160.03	10.24

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

其月度收益率见表 3－3：

表 3－3

	上证综指	浦发银行
2012－12－3	0. 145956	0. 356522
2013－1－4	0. 051249	0. 166667
2013－2－1	－0. 00831	－0. 03755
2013－3－1	－0. 05452	－0. 08944
2013－4－1	－0. 02625	－0. 02612
2013－5－2	0. 056329	0. 064378
2013－6－3	－0. 13969	－0. 16532
2013－7－1	0. 007372	－0. 05072
2013－8－1	0. 052453	0. 143766
2013－9－2	0. 02938	0. 139043

我们可以绘出以上月收益率的散点图：

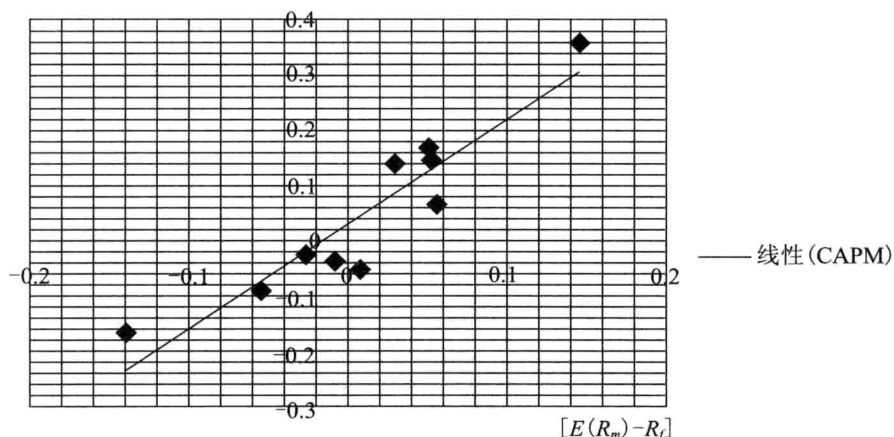


图 3. 6

利用 EViews 计量软件做回归分析，得到：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

Dependent Variable：RI_RF
Method：Least Squares
Date：10/25/13 Time：22：00
Sample：1 10
Included observations：10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0. 028783	0. 019295	1. 491742	0. 1741
RM_RF	1. 872483	0. 262530	7. 132463	0. 0001
R-squared	0. 864112	Mean dependent var		0. 050122
Adjusted R-squared	0. 847126	S. D. dependent var		0. 154166
S. E. of regression	0. 060278	Akaike info criterion		-2. 602854
Sum squared resid	0. 029067	Schwarz criterion		-2. 542337
Log likelihood	15. 01427	F-statistic		50. 87203
Durbin-Watson stat	2. 493524	Prob（F-statistic）		0. 000099

由此可见，浦发银行的月度风险溢价与市场组合的风险溢价成正比，β 系数的估计值为 1. 872483，t 统计量通过了显著性检验，β 值不为零。而系数 α 的估计值为 0. 028783，t 统计量为 1. 491742，未通过显著性检验，α 值为零。从整体上来看，该方程具有较高的拟合效果，调整后的 R 平方为 0. 847。这说明，CAPM 基本成立。

去掉常数项 α 后，我们可以得到如下较为理想的结果：

Dependent Variable：RI_RF
Method：Least Squares
Date：10/25/13 Time：22：15
Sample：1 10
Included observations：10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM_RF	1. 933209	0. 276446	6. 993076	0. 0001
R-squared	0. 826313	Mean dependent var		0. 050122
Adjusted R-squared	0. 826313	S. D. dependent var		0. 154166
S. E. of regression	0. 064250	Akaike info criterion		-2. 557431
Sum squared resid	0. 037153	Schwarz criterion		-2. 527172
Log likelihood	13. 78715	Durbin-Watson stat		2. 110943

当然，这个结果可能还有需要改进的地方。此外，需要说明的是，我们这里做的是一个对 CAPM 模型的简单回归分析，从严格意义上讲，上面的简单回归分析不完全服从古典假设，尤其是上证综指的计算中包含了浦发银行。解决这个问题，就需要使用工具变量法或其它方法。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第四章

最小二乘法与多因素模型
——含多个自变量的最小二乘法

在这一章以前，我们讨论的都是含有一个自变量的用最小二乘法所得到的拟合直线，即回归模型为 $y_i = a + bx_i + \varepsilon_i$ 。但是，在现实生活中，引起因变量变化的往往并不仅是一个自变量。从现在起我们就来讨论两个自变量变化对因变量的影响，这种情况在现实中俯拾皆是。比如：经济增长往往就受到资本性支出与经常性支出两种因素影响。同一元回归分析类似，我们这里所用的回归模型为：$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$，其中 $i = 1,2,\cdots,n$，这里 n 表示每个变量样本的个数，拟合直线仍可写成 $y_i = \hat{y}_i + \hat{\varepsilon}_i$ 。不过，这次 $\hat{y}_i$ 的值依赖于 x_{1i} 和 x_{2i} 这两个自变量。

为了考察由两个自变量对因变量所产生的共同影响，仍可以用最小二乘法来拟合出它们的变化趋势。这和我们推导一元回归模型的思路是一样的：使拟合值尽可能地接近真实值。上述思路可以通过使得二者之间的残差平方和最小来实现。

第一节　含两个自变量的最小二乘法

为了使各实际点到拟合直线 $\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$ 的垂直距离最小，我们可考察以下极值问题：

$$\min_{\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2} \sum (y_i - \hat{y}_i)^2 = \min_{\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2} \sum [y_i - (\hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i})]^2$$

根据极值条件可得：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\begin{cases} \dfrac{\partial \sum [y_i - (\hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i})]^2}{\partial \hat{\alpha}} = 0 \\ \dfrac{\partial \sum [y_i - (\hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i})]^2}{\partial \hat{\beta}_1} = 0 \\ \dfrac{\partial \sum [y_i - (\hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i})]^2}{\partial \hat{\beta}_2} = 0 \end{cases}$$

由于求解过程繁琐，这里暂且略去，仅给出结果如下：

$$\hat{\alpha} = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$$

$$\hat{\beta}_1 = \frac{\sum \dot{x}_{2i}^2 \sum \dot{x}_{1i}\dot{y}_i - \sum \dot{x}_{1i}\dot{x}_{2i} \sum \dot{x}_{2i}\dot{y}_i}{(\sum \dot{x}_{1i}^2)(\sum \dot{x}_{2i}^2) - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$$

$$\hat{\beta}_2 = \frac{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}\dot{y}_i - \sum \dot{x}_{1i}\dot{x}_{2i} \sum \dot{x}_{1i}\dot{y}_i}{(\sum \dot{x}_{1i}^2)(\sum \dot{x}_{2i}^2) - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$$

其中 $\dot{x}_{1i} = x_{1i} - \bar{x}_1$，$\dot{x}_{2i} = x_{2i} - \bar{x}_2$，$\dot{y}_i = y_i - \bar{y}$。

$$\sum \dot{x}_{1i}\dot{x}_{2i} = \sum x_{1i}x_{2i} - n\dot{x}_1\dot{x}_2$$

$$\sum \dot{x}_{1i}\dot{y}_i = \sum x_{1i}y_i - n\bar{x}_1\bar{y}$$

$$\sum \dot{x}_{2i}\dot{y}_i = \sum x_{2i}y_i - n\bar{x}_2\bar{y}$$

通过以上的分析，我们对二元回归会有比以前更为清楚的理解。同时，我们需要指出，在含两个自变量的回归模型中，*ESS*、*TSS*、*RSS* 的定义有效，只不过 $\hat{y}$ 的形式稍作变化罢了。

$$ESS = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})^2$$

$$TSS = \sum (y_i - \bar{y})^2$$

$$RSS = \sum (\hat{y}_i - \bar{y})^2 = \sum (\hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} - \bar{y})^2$$

$$R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS}$$

二元回归所得到的拟合直线的性质与一元回归所得到的拟合直线的性质一样，也就是说，在二元回归情况下，拟合直线同样具有如下性质：

1. $\sum \hat{\varepsilon}_i = 0$；

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

2. $\bar{y} = \bar{\hat{y}}$；

3. $\text{cov}(\hat{\varepsilon}_i, x_{1i}) = \text{cov}(\hat{\varepsilon}_i, x_{2i}) = 0$；

4. $\text{cov}(\hat{\varepsilon}_i, \hat{y}_i) = 0$。

我们仍有必要再重复一下，$\hat{\varepsilon}_i$ 是样本的残差，它等于 $y_i - \hat{y}_i$，而 ε_i 则指整个母体的残差，在应用中应加倍注意。

第二节　二元回归最小二乘法解法又探

这里，我们先来看看 $\hat{\beta}_1$，这个等式究竟能告诉我们什么？从前面的推导中我们已经得到了：

$$\hat{\beta}_1 = \frac{\sum \dot{x}_{2i}^2 \sum \dot{x}_{1i}\dot{y}_i - \sum \dot{x}_{1i}\dot{x}_{2i} \sum \dot{x}_{2i}\dot{y}_i}{(\sum \dot{x}_{1i}^2)(\sum \dot{x}_{2i}^2) - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$$

现在仔细研究一下：

$$\sum \dot{x}_{1i}\dot{x}_{2i} = \sum (\dot{x}_{1i} - \bar{x}_1)(\dot{x}_{2i} - \bar{x}_2) = n \cdot \text{cov}(x_{1i}, x_{2i})$$

有了 $\text{cov}(x_{1i}, x_{2i})$，自然会令人紧接着想起 $\rho_{x_{1i},x_{2i}}$。一般情况下，这两个指标总是联系在一起的。将 $\rho_{x_{1i},x_{2i}}$ 平方，可得

$$\rho^2_{x_{1i},x_{2i}} = \frac{[\sum (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)]^2}{\sum (x_{1i} - \bar{x}_1)^2 \sum (x_{2i} - \bar{x}_2)^2} = \frac{[\sum \dot{x}_{1i}\dot{x}_{2i}]^2}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2}$$

有了以上的“根基”，我们可以对 $\hat{\beta}_1$“动手术”了。先从分母入手：提出 $\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2$，使 $\hat{\beta}_1$ 变成

$$\hat{\beta}_1 = \frac{(\sum \dot{x}_{1i}\dot{y}_i)\sum \dot{x}_{2i}^2 - n \cdot \text{cov}(x_{1i}, x_{2i})(\sum \dot{x}_{2i}\dot{y}_i)}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 [1 - \frac{(\sum \dot{x}_{1i}\dot{x}_{2i})^2}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2}]}$$

由于 $\frac{(\sum \dot{x}_{1i}\dot{x}_{2i})^2}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2} = \rho^2_{x_{1i},x_{2i}}$，$\hat{\beta}_1$ 可进一步写成

$$\hat{\beta}_1 = \frac{(\sum \dot{x}_{1i}\dot{y}_i)\sum \dot{x}_{2i}^2 - n \cdot \text{cov}(x_{1i}, x_{2i})(\sum \dot{x}_{2i}\dot{y}_i)}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 [1 - \rho^2_{x_{1i},x_{2i}}]}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

对 $\hat{\beta}_1$ 做变形，目的是为了更清楚地看到 $\hat{\beta}_1$ 所反映的是 x_{1i} 本身变化对 y_i 的净影响。

如果在实际中确实是有两种变量因素同时对 y 发生影响，而我们试图用一元而不是二元回归变量来回归时，将得到 $\hat{y}_i = \hat{\alpha} + \hat{\beta}_1^* x_{1i}$，这里的 $\hat{\beta}_1^* = \frac{\sum (x_{1i} - \bar{x}_1)(y_i - \bar{y})}{\sum (x_{1i} - \bar{x}_1)^2} = \frac{\sum \dot{x}_{1i}\dot{y}_i}{\sum \dot{x}_{1i}^2}$。

再来考察一下二元回归模型 $\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$ 所得到的 $\hat{\beta}_1$：

$$\hat{\beta}_1 = \frac{(\sum \dot{x}_{1i}\dot{y}_i)\sum \dot{x}_{2i}^2 - n \cdot \mathrm{cov}(x_{1i}, x_{2i})(\sum \dot{x}_{2i}\dot{y}_i)}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 [1 - \rho_{x_{1i},x_{2i}}^2]}$$

这里，我们根据 $\rho_{x_{1i},x_{2i}}^2$ 的不同情况阐述 $\hat{\beta}_1$ 与 $\hat{\beta}_1^*$ 之间的相互关系：

当 $\rho_{x_{1i},x_{2i}}^2 = 0$, $\mathrm{cov}(x_{1i}, x_{2i}) = 0$ 时, $\hat{\beta}_1 = \frac{(\sum \dot{x}_{1i}\dot{y}_i)\sum \dot{x}_{2i}^2}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2} = \frac{\sum \dot{x}_{1i}\dot{y}_i}{\sum \dot{x}_{1i}^2} = \hat{\beta}_1^*$。

当 $\rho_{x_{1i},x_{2i}}^2 = 1$ 时, $\hat{\beta}_1$ 将会趋近于无穷大，从而使 $\hat{\beta}_1$ 无法求出。这就是所谓的多重共线性（Multicollinearity）问题。

当 $0 < \rho_{x_{1i},x_{2i}}^2 < 1$ 时，无法判断 $\hat{\beta}_1, \hat{\beta}_1^*$ 谁大，因为这将取决于 $\mathrm{cov}(x_{1i}, x_{2i})$，$\sum \dot{x}_{2i}\dot{y}_i$ 的符号和 $\sum \dot{x}_{2i}^2$ 的大小。

那么探讨 $\hat{\beta}_1$ 和 $\hat{\beta}_1^*$ 的相互关系的意义究竟何在呢？关键在于，在两个自变量共同作用下，x_{1i}、x_{2i} 对 y_i 的影响将来自四个方面：

1. x_{1i} 本身变化对 y_i 的净作用；
2. x_{1i} 的变化引起 x_{2i} 的相应变化，通过 x_{2i} 作用于 y_i；
3. x_{2i} 本身变化对 y_i 的净作用；
4. x_{2i} 的变化引起 x_{1i} 的相应变化，通过 x_{1i} 作用于 y_i。

具体过程如图 4.1 所示。

复杂的表现在于，阴影箭头是 x_{1i}（或 x_{2i}）综合的对 y_i 的作用。在 x_{1i} 这个作用中，既包括 x_{1i} 本身对 y_i 的净影响（实线箭头），也包括了由 x_{2i} 变化引起的 x_{1i} 的变动对 y_i 的影响（虚线箭头）。从 x_{2i} 这一方面来说，既包括 x_{2i} 本身变化对 y_i 的净影响（实线箭头），也包括了由 x_{1i} 变化引起 x_{2i} 的变化而作用于 y_i 的影响（虚线箭头）。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

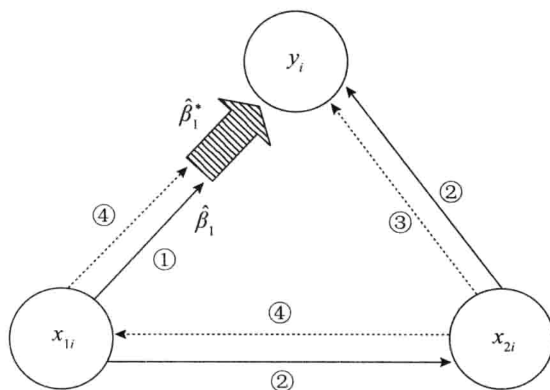


图 4.1

现在单看 x_{1i}。我们所求的 $\hat{\beta}_1 = \dfrac{\sum \dot{x}_{1i}^2 \sum \dot{x}_{1i}\dot{y}_i - \sum \dot{x}_{1i} \sum \dot{x}_{2i}\dot{y}_i}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$，实际描述的是 x_{1i} 本身运动变化对 y_i 所产生的净效应，如图 4.1 所示；而 $\hat{\beta}_1^*$ 则描述的是 x_{1i} 本身对 y_i 的净影响和由 x_{2i} 变化所引起的 x_{1i} 的变化对 y_i 的影响这二者之和，即用箭头①和④表示的影响。

再来看 $\hat{\beta}_1$，它是如何消去 x_{2i} 对 x_{1i} 的影响的呢？

$$\hat{\beta}_1 = \frac{(\sum \dot{x}_{1i}\dot{y}_i) \sum \dot{x}_{2i}^2 - n \cdot \mathrm{cov}(x_{1i}, x_{2i})(\sum \dot{x}_{2i}\dot{y}_i)}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 [1 - \rho_{x_{1i}, x_{2i}}^2]}$$

虽然 $\hat{\beta}_1$ 中有 x_{2i}，但是它利用了 $\mathrm{cov}(x_{1i}, x_{2i})$ 和 $\rho_{x_{1i}, x_{2i}}^2$，这两个减项消除了 x_{2i} 对 y_i 的间接影响，而余下 x_{1i} 本身对 y_i 的净影响。

以上的设想还不能得到令人信服的结论，现在我们就来实际论证一下我们的观点。我们用消除残差法考察 $\hat{\beta}_1$ 与对 y_i 的净影响。

在 $y_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \varepsilon_i$ 回归模型中，x_{1i}、x_{2i} 共同作用于 y_i。在回归模型 $y_i = \hat{\alpha} + \hat{\beta}_1^* x_{1i} + \varepsilon_i$ 中，$\hat{\beta}_1^*$ 实际是聚合了 x_{1i} 本身和 x_{2i} 通过 x_{1i} 对 y_i 的影响。

如果想求出 x_{1i} 本身变化对 y_i 的影响，简单的办法就是从 x_{1i} 的总体变化中减去 x_{2i} 变化所引起的 x_{1i} 变化对 y_i 的影响。看一看图 4.2 就会明白这条思路了。

第一步，先求 x_{1i} 和 x_{2i} 的关系，拟合出 x_{2i} 对 x_{1i} 影响的方程。这个方程的因变量是 x_{1i}，自变量是 x_{2i}，是一个一元回归方程，对已知的 x_{1i}、x_{2i} 求出拟合方程，不妨用 $\hat{x}_{1i} = \hat{c} + \hat{d}x_{2i}$ 表示。依此可知：$\hat{c} = \bar{x}_1 - \hat{d}\bar{x}_2$，$\hat{d} = \dfrac{\sum \dot{x}_{1i}\dot{x}_{2i}}{\sum \dot{x}_{2i}^2}$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

图 4.2

第二步，从 x_{1i} 总体变动中剔除掉由 x_{2i} 变化所引起的 x_{1i} 变化部分，可以引进一个变量 V_{1i}：$V_{1i} = x_{1i} - \hat{x}_{1i} = x_{1i} - \hat{c} - \hat{d}x_{2i}$。

注意，这里的 x_{1i} 乃是 $\hat{\beta}_1^*$ 下的 x_{1i}，也就是说 x_{1i} 自身及 x_{2i} 影响的综合体，V_{1i} 是什么呢？我们知道，在回归模型 $x_{1i} = c + dx_{2i} + \varepsilon_i$ 中，存在一个残差，所以从表面上看 V_{1i} 表示拟合过程中的残差。同时，由上式可知，V_{1i} 实际上已表示单纯 x_{1i} 自身变化，而将用 V_{1i} 对 y_i 进行第二次拟合，第二次拟合中的残差将把第一次拟合中的真正残差包括进去，因此，不妨认为 V_{1i} 就是 x_{1i} 自身变化量。

第三步，考察在消除了 x_{2i} 的干扰后的 V_{1i} 即 x_{1i} 自身变化对 y_i 的作用。我们可以用新方程形式 $y_i = f + gV_{1i} + \varepsilon_i$ 表示对 V_{1i} 和 y_i 进行的拟合结果：$\hat{y}_i = \hat{f} + \hat{g}V_{1i}$。

在求这个拟合方程的拟合值之前，再看一看 V_{1i}。已知 $V_{1i} = x_{1i} - \hat{x}_{1i}$，撇开具体意义不谈，单从式子本身看，我们已经熟悉 $x_{1i} - \hat{x}_{1i}$，表示真实值 x_{1i} 到拟合值 $\hat{x}_{1i}$ 的差距，那么 V_{1i} 本身成为一个一元线性回归中的残差值了（注意 V_{1i} 的双重作用，实际上 V_{1i} 本身就已是个残差，它本身又包括一个真正的残差，你最好在这儿停下来想想 V_{1i} 究竟有哪些身份，特殊在哪里）。从抽象理论讲，V_{1i} 作为 $x_{1i} - \hat{x}_{1i}$ 的残差，应该具有一元线性回归拟合残差的性质。在一元线性回归模型 $y_i = \alpha + \beta x_i + \varepsilon_i$ 中，拟合残差 $\hat{\varepsilon}_i$ 是有如下性质的：$\sum \hat{\varepsilon}_i = 0$，$\bar{\varepsilon}_i = 0$，那么，也就顺理成章的有 $\sum V_{1i} = 0$，$\bar{V}_{1i} = 0$。

有了这些性质，就可以揭开 $\hat{g}$ 的真实面目了，对于一元线性回归模型：$y_i = f + gV_{1i} + \varepsilon_i$，按一元线性回归方程对拟合直线的求法，我们自然可以得到：$\hat{g} = \dfrac{\sum v_{1i}\dot{y}_i}{\sum v_{1i}^2}$，其中 $v_{1i} =$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$V_{1i}-\overline{V}_{1i}$。

先看上式分子：$\sum v_{1i}\dot{y}_i$。$\sum v_{1i}\dot{y}_i$ 实际上等于 $\sum V_{1i}\dot{y}_i$，这是因为 $\overline{V}_{1i}=0$，$v_{1i}=V_{1i}-\overline{V}_{1i}=V_{1i}$，故而 $\sum v_{1i}\dot{y}_i=\sum V_{1i}\dot{y}_i$。从 $\sum v_{1i}\dot{y}_i$ 动手：

$$\begin{aligned}\sum v_{1i}\dot{y}_i &= \sum V_{1i}\dot{y}_i \\ &= \sum (x_{1i}-\hat{x}_{1i})\dot{y}_i \\ &= \sum (x_{1i}-\hat{c}-\hat{d}x_{2i})\dot{y}_i \\ &= \sum (x_{1i}-(\overline{x}_1-\hat{d}\overline{x}_2)-\hat{d}x_{2i})\dot{y}_i\end{aligned}$$

（把 $\hat{c}=\overline{x}_1-\hat{d}\overline{x}_2$ 代进去，用 $\dot{x}_{1i}=x_{1i}-\overline{x}_1$ 和 $\dot{x}_{2i}=x_{2i}-\overline{x}_2$ 代替原式）

$$= \sum (\dot{x}_{1i}-\hat{d}\dot{x}_{2i})\dot{y}_i$$

由于 $\hat{d}=\dfrac{\sum x_{1i}x_{2i}-n\overline{x}_1\overline{x}_2}{\sum (x_{2i}-\overline{x}_2)^2}=\dfrac{\sum \dot{x}_{1i}\dot{x}_{2i}}{\sum \dot{x}_{2i}^2}$，$\sum v_{1i}\dot{y}_i$ 可以进一步写成：

$$\sum\left(\dot{x}_{1i}-\frac{\sum \dot{x}_{1i}\dot{x}_{2i}}{\sum \dot{x}_{2i}^2}\cdot\dot{x}_{2i}\right)\dot{y}_i=\frac{\sum \dot{x}_{1i}\dot{y}_i\sum \dot{x}_{2i}^2-\sum \dot{x}_{1i}\dot{x}_{2i}\sum \dot{x}_{2i}\dot{y}_i}{\sum \dot{x}_{2i}^2}$$

再分析分母：

$$\begin{aligned}\sum v_{1i}^2 &= \sum V_{1i}^2 \\ &= \sum (x_{1i}-\hat{x}_{1i})^2 \\ &= \sum (x_{1i}-\hat{c}-\hat{d}x_{2i})^2 \\ &= \sum [x_{1i}-(\overline{x}_1-\hat{d}\overline{x}_2)-\hat{d}x_{2i}]^2 \\ &= \sum (\dot{x}_{1i}-\hat{d}\dot{x}_{2i})^2 \\ &= \sum \dot{x}_{1i}^2-2\left(\sum \dot{x}_{1i}\dot{x}_{2i}\right)\frac{\sum \dot{x}_{1i}\dot{x}_{2i}}{\sum \dot{x}_{2i}^2}+\left(\sum \dot{x}_{2i}^2\right)\left(\frac{\sum \dot{x}_{1i}\dot{x}_{2i}}{\sum \dot{x}_{2i}^2}\right)^2 \\ &= \frac{\sum \dot{x}_{1i}^2\sum \dot{x}_{2i}^2+\left(\sum \dot{x}_{1i}\dot{x}_{2i}\right)^2-2\left(\sum \dot{x}_{1i}\dot{x}_{2i}\right)^2}{\sum \dot{x}_{2i}^2} \\ &= \frac{\sum \dot{x}_{1i}^2\sum \dot{x}_{2i}^2-\left(\sum \dot{x}_{1i}\dot{x}_{2i}\right)^2}{\sum \dot{x}_{2i}^2}\end{aligned}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

最后把分子分母相除：

$$\hat{g}=\frac{\dfrac{\sum\dot{x}_{1i}\dot{y}_i\sum\dot{x}_{2i}^2-\sum\dot{x}_{1i}\dot{x}_{2i}\sum\dot{x}_{2i}\dot{y}_i}{\sum\dot{x}_{2i}^2}}{\dfrac{\sum\dot{x}_{1i}^2\sum\dot{x}_{2i}^2-\left(\sum\dot{x}_{1i}\dot{x}_{2i}\right)^2}{\sum\dot{x}_{2i}^2}}$$

上面所求的 $\hat{g}$ 是我们从 $y_i=\hat{\alpha}+\hat{\beta}_1^*+x_{1i}+\varepsilon_i$ 中剔掉了 x_{2i} 的干扰后，求出 x_{1i} 对 y_i 的影响的净作用，它实际上等于我们二元回归模型：$\hat{y}_i=\hat{\alpha}+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}$ 中的 $\hat{\beta}_1$ 的值。这也进一步证明 $\hat{\beta}_1$ 就是表示 x_{1i} 本身变化对 y_i 的净影响。

第三节　从另一个角度看估计系数

记得我们在第二节“最小二乘法又探”中得出的结论吗？当 $\mathrm{cov}(x_{1i},x_{2i})=0$，$\rho_{x_{1i},x_{2i}}=0$ 时，将有 $\hat{\beta}_1=\hat{\beta}_1^*$。这个结论有助于我们揭示 $\hat{\beta}_1^*$ 的本性。

当 $\mathrm{cov}(x_{1i},x_{2i})=0$ 时，意味着如果 x_{1i} 改变了一个单位，x_{2i} 却不会变动，这是因为它们不相关。若用式子表述上面的想法，就是：旧的 $\hat{y}_i=\hat{\alpha}+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}$ 在 x_{1i} 改变一个单位，设为 1，就有新的 $\hat{y}_i=\hat{\alpha}+\hat{\beta}_1(x_{1i}+1)+\hat{\beta}_2x_{2i}$ 出现。那么：$\hat{y}_i(\text{新})-\hat{y}_i(\text{旧})=\hat{\beta}_1$。

因此，$\hat{\beta}_1=\hat{y}_i(\text{新})-\hat{y}_i(\text{旧})=\Delta y$。

这说明了 $\hat{\beta}_1$ 在 $\mathrm{cov}(x_{1i},x_{2i})=0$ 时等于 x_{1i} 增加一单位后对 y 的影响，也就是 y 的增量。

当 $\mathrm{cov}(x_{1i},x_{2i})\neq 0$ 时，假设 x_{1i}、x_{2i} 的关系为：$x_{2i}=c+dx_{1i}+\eta$。

如果 x_{1i} 变化了 Δx_{1i} 个单位，将引起 x_{2i} 变化 $d\Delta x_{1i}$ 单位，即

$$\hat{y}_i(\text{旧})=\hat{\alpha}+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}$$

$$\hat{y}_i(\text{新})=\hat{\alpha}+\hat{\beta}_1(x_{1i}+\Delta x_{1i})+\hat{\beta}_2(x_{2i}+d\Delta x_{1i})$$

那么：$\hat{y}_i(\text{新})-\hat{y}_i(\text{旧})=(\hat{\beta}_1+\hat{\beta}_2d)\Delta x_{1i}$，这也就是说，$\Delta y=\hat{\beta}_1\Delta x_{1i}+(\hat{\beta}_2d)\Delta x_{1i}$。

从上面结果可以看出，在 Δy 中，$\hat{\beta}_1\Delta x_{1i}$ 代表直接作用，而 $(\hat{\beta}_2d)\Delta x_{1i}$ 则表示 x_{1i} 通过 x_{2i} 对 y_i 发挥的间接影响——这一点也可以直观看到：表示 x_{2i} 影响的 $\hat{\beta}_2$ 在该项中出现了。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第四节 多重共线性

多重共线性（Multicollinearity）是指两个自变量之间存在完全的线性相关关系的情况。在二元回归中，也就是指在两个自变量 x_{1i} 、x_{2i} 之间存在完全线性相关关系：$x_{1i} = c + dx_{2i}$ 。注意：这里不是 $x_{1i} = c + dx_{2i} + \varepsilon_i$，当 x_{2i} 变化一个单位时，在 $x_{1i} = c + dx_{2i}$ 式子中的 x_{1i} 变化的是 dx_{2i} 个单位。而在 $x_{1i} = c + dx_{2i} + \varepsilon_i$ 变化 $dx_{2i} + \varepsilon_i$ 个单位。这是因为当 x_{1i} 、x_{2i} 完全线性相关时，$\rho^2_{x_{1i},x_{2i}} = 1$ 。

在这种情况下，估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 都将趋于无穷大。因为当两个自变量完全线性相关时，二元回归模型相当于对同一变量回归了两次。解决这一问题时，会想到把 $x_{1i} = c + dx_{2i}$ 代入到二元回归模型 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ 中去，只对 x_{2i} 求回归，即：

$$y_i = \alpha + \beta_1(c + dx_{2i}) + \beta_2 x_{2i} + \varepsilon_i = (\alpha + \beta_1 c) + (\beta_1 d + \beta_2)x_{2i} + \varepsilon_i$$

令 $\alpha^* = \alpha + \beta_1 c$，$\beta^* = \beta_1 d + \beta_2$，则 $y_i = \alpha^* + \beta^* x_{2i} + \varepsilon_i$ 。

这是一个新的一元线性回归方程，并且在回归后可以得到估计量 $\hat{\alpha}^*$ 和 $\hat{\beta}^*$ 。由于 $\alpha^* = \alpha + \beta_1 c$，$\beta^* = \beta_1 d + \beta_2$，$\hat{\alpha}$，$\hat{\beta}_1$，$\hat{\beta}_2$ 可以从方程组 $\begin{cases}\alpha^* = \alpha + \beta_1 c \\ \beta^* = \beta_1 d + \beta_2\end{cases}$ 中得到。

在这个方程组中，c 、d 是已知的，$\hat{\alpha}^*$ 、$\hat{\beta}^*$ 也可求到，但是 α 、β_1 、β_2 这三个未知变量却只含于两个方程中，我们无法得到未知量的值。这也告诉我们，在完全线性相关时不能轻易地使用最小二乘法直接对二元线性回归模型进行回归。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第五节　一般线性回归

至今为止，我们已经讨论了两种情况的线性回归模型。第一种情况是一元线性回归模型，它体现了 x 对 y 的影响。而没有考察除 x 外的其它因素对 y 的作用。二元线性回归模型是第二种情况，表示两因素 x_{1i} 和 x_{2i} 对 y 的影响。实际上，在现实中这仍是一个特殊的情况。幸运的是，最小二乘法可以扩展到由一系列自变量影响因变量的情形下。这样，我们就可以认为，一般线性回归模型为：

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \varepsilon_i = \hat{y}_i + \hat{\varepsilon}_i$$

其中，$\hat{y} = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki} \qquad (i = 1,2,\cdots,n)$。

这里有关最小二乘法的性质仍然适用，于是 RSS、ESS、TSS 的定义范围也得到了扩展。我们仍可以证明 $\frac{ESS}{TSS} + \frac{RSS}{TSS} = 1$ 在这种情况下也是成立。换言之，$R^2 = 1 - \frac{ESS}{TSS}$ 也是成立的。

对一般线性回归模型的处理，多采取矩阵的表示方法，具体如下：

我们可以将模型：

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \varepsilon_i$$

具体化为：

$$y_1 = \alpha + \beta_1 x_{11} + \beta_2 x_{21} + \beta_3 x_{31} + \cdots + \beta_k x_{k1} + \varepsilon_1$$

$$y_2 = \alpha + \beta_1 x_{12} + \beta_2 x_{22} + \beta_3 x_{32} + \cdots + \beta_k x_{k2} + \varepsilon_2$$

……

$$y_n = \alpha + \beta_1 x_{1n} + \beta_2 x_{2n} + \beta_3 x_{3n} + \cdots + \beta_k x_{kn} + \varepsilon_n$$

我们可以用矩阵将上述表达式表示为：

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} \cdots x_{k1} \\ 1 & x_{12} & x_{22} \cdots x_{k2} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \cdots x_{kn} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

令：$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & x_{21} \cdots x_{k1} \\ 1 & x_{12} & x_{22} \cdots x_{k2} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \cdots x_{kn} \end{bmatrix}$, $\beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$, $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$

则：$Y = X\beta + \varepsilon$

我们可以仿效一元线性回归模型中求 $\hat{\beta}$ 的最小二乘法，求出 β 的估计值。也就是说，求出这样的 β，使得 $\varepsilon'\varepsilon$。

$$\begin{aligned} \underset{\beta}{\mathrm{Min}}\varepsilon'\varepsilon &= (Y - X\beta)'(Y - X\beta) \\ &= (Y' - \beta'X')(Y - X\beta) \\ &= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta \\ &= Y'Y - 2\beta'X'Y + \beta'X'X\beta \end{aligned}$$

对上式求 β 的一阶导数：$\dfrac{\partial\ [Y'Y - 2\beta'X'Y + \beta'X'X\beta]}{\partial\ \beta} = -2X'Y + 2X'X\beta$。

由于 $X \neq 0$，X' 可逆，我们可以得到：$\hat{\beta} = (X'X)^{-1}X'Y$。

这里的 β，很像一元回归中的 $\hat{\beta} = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$，其中，$(X'X)^{-1}$ 相当于 $\sum(x_i - \bar{x})^2$，$X'Y$ 相当于 $\sum(x_i - \bar{x})(y_i - \bar{y})$。

同理，我们还可以得到一般线性回归的 ESS、TSS、RSS 和 R^2。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第五章 古典线性回归模型

我们在此之前所进行的简单回归，主要是对实际中的一些现象（点），通过回归找出能够反映这些现象的变化规律或趋势的一条直线。简单地说，简单回归就是由点得到线的过程。作为真正的计量分析，还要注意到在经济世界中，本身存在着的经济规律的实际作用。这些经济规律的作用，通过现实经济生活中一些复杂现象表现出来。因此，这些现象既有某种确定性，又具有某种不确定性。计量分析所进行的回归，是在假设这些数据背后有一个数据产生过程，即现象后面有经济规律的作用，而力图通过回归模型来反映这一规律。这是简单回归与计量分析的不同之处。

在运用最小二乘法进行线性回归时，最小二乘法保证了如下条件的成立：$\sum \hat{\varepsilon}_i = 0$，$\sum \hat{\varepsilon}_i x_i = 0$，$\sum \hat{\varepsilon}_i \hat{y}_i = 0$。而实际上，数据并不一定使这些条件都成立，也即现实中这些条件并不一定得到满足。我们不妨假设我们所接触到的数据都有一个数据产生过程，即通过如下方程产生出来的一些离散点：

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \varepsilon_i \quad (i = 1,2,\cdots,n)$$

现在的问题是，假设我们现在仅知道一些离散点，也知道这些离散点产生的背后存在着上述方程所表示的数据产生过程。这是我们进行一切工作的前提假设。在讨论数据产生过程具体形式之前，必须先对整个 ε_i 进行假设。在前面的简单回归中，我们运用最小二乘法，曾对已知离散点中的随机项 $\hat{\varepsilon}_i$，保证过 $\sum \hat{\varepsilon}_i = 0$，$\sum \hat{\varepsilon}_i x_i = 0$，$\sum \hat{\varepsilon}_i \hat{y}_i = 0$。$\hat{\varepsilon}_i$ 表示的是一个样本的残差，从样本和母体的关系，我们自然可以得出 $\hat{\varepsilon}_i$ 是 ε_i 的一个子集：$\hat{\varepsilon}_i$ 是样本离散点的随机扰动项，而 ε_i 为整个母体的随机扰动项。样本一般来说总会反映出母体的一些性质，因此，我们可以假设整个母体的随机扰动项 ε_i 也满足类似性质：$\sum \varepsilon_i = 0$，$\sum \varepsilon_i x_i = 0$，$\sum \varepsilon_i y_i = 0$。由于 ε_i 在产生数据过程中起了很大作用，故而只有在研究了 ε_i 的产生规律前提下，才能进一步探讨数据产生规律的一些性质。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第一节　有关 ε_i 的古典模型假设

假设 5.1.1　随机扰动项 ε_i 垂直波动。

这个假设是说，样本数据点只沿着 y_i 的方向随机地在真实直线附近垂直跳动，也就是说，这种波动是围绕真实直线上下波动。

运用图形会得到很清晰的解释。先看图 5.1，这是符合假设 5.1.1 的条件的情况。图 5.1 中，真实直线用——表示，$y_i = 2 - 1.2x_i + \varepsilon_i$ 表示的是真正的数据产生过程。

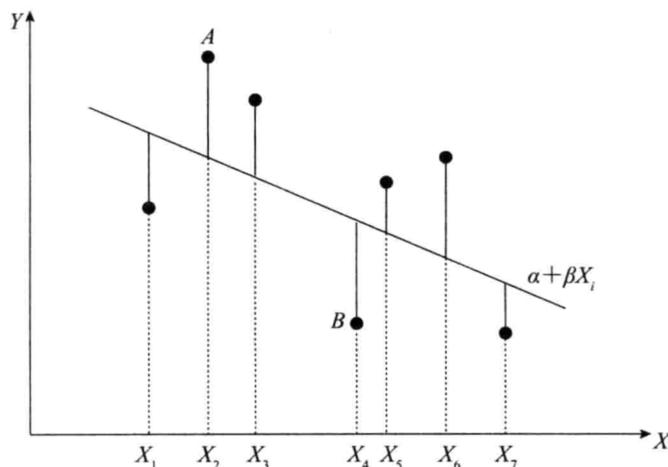


图 5.1

由于扰动项的作用，在实际中表现为：对应于 x_2，x_4 的值的点为 A、B 两点。这些点要么在真实直线上方，要么在真实直线下方。但无论如何，对这些点而言，对应于每一个实际的 x_i，它们总是垂直变动，没有横向的偏移。这实际上也就是假设我们所观察到的 x_i 是准确无误的，就是实际中的真实 x_i 没有丝毫偏差，而对应于 x_i 的 y_i 却存在着偏差。

再看图 5.2，它表现了在数据产生过程中不仅 y_i 上下波动，x_i 也左右波动，即对同一个数据是由两个随机扰动项的作用下加以影响后产生的。图 5.2 中用——表示真实直线，真实的数据产生过程乃是 $y_i = 2 - 1.2(x_i + \eta_i) + \varepsilon_i$，于是很可能对于一个实际的 x_i，由于其扰动项 η_i 的作用，使其反映到我们的观察值时发生了偏差，变成 $x_i + \eta_i$。在 $x_i + \eta_i$ 基础

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



上的 y_i，再一次由于扰动项 ε_i 的作用而发生偏离。这类数据的生成过程叫变量误差模型（Errors in Variable Model）。很显然，这种情况不符合古典假设 5. 1. 1。

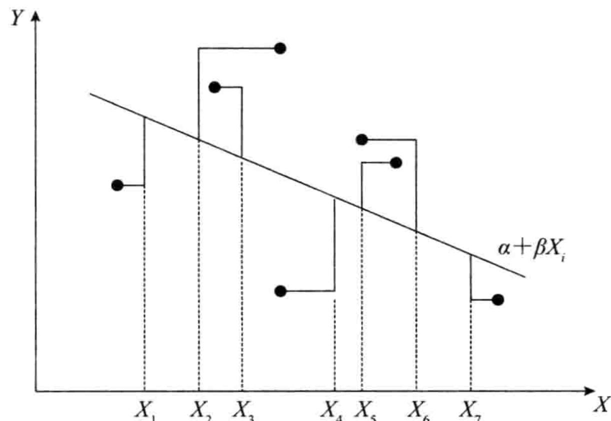


图 5. 2

假设 5. 1. 2　残差分布均值为零，即偏离程度的均值为 0，也即：$E(\varepsilon_i)=0\ (i=1,2,\cdots,n)$。

必须注意：$E(\hat{\varepsilon}_i)=0$ 是最小二乘法保证了的，也就是说，只要使用最小二乘法，就一定会有 $E(\hat{\varepsilon}_i)=0$，即样本残差的均值为零。而 $E(\varepsilon_i)=0$ 则是一个假设，我们假设总体残差 ε_i 的期望为零，即总体残差 ε_i 从整体上说对回归的影响不大。

假设 5. 1. 3　方差不变，即：

$$Var(\varepsilon_i)=E(\varepsilon_i^2)=\sigma^2\ (i=1,2,\cdots,n)$$

这条假设说明对所有的 ε_i，其方差 σ^2 不随其大小差异不同而变化，这种情况称为同方差（Homoskedasticity）（见图 5. 3）。如果 $Var(\varepsilon_i)=E[\varepsilon_i-E(\varepsilon_i)]^2=E(\varepsilon_i^2)=\sigma_i^2$，即 σ_i^2 为一个变量，而非一个定值，这种情况就叫异方差（Heteroskedasticity）（见图 5. 4）。

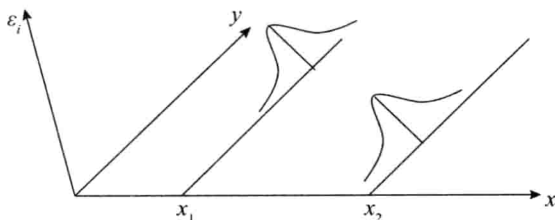


图 5. 3

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

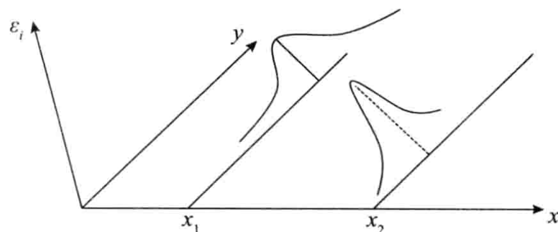


图 5.4

假设 5.1.4　偏差相互独立，即：$\varepsilon_i,\varepsilon_j$ 不相关，也就是说，对于所有 $i\neq j$，有 $E(\varepsilon_i,\varepsilon_j)=0$。

$E(\varepsilon_i,\varepsilon_j)$ 怎么来的呢？我们从假设 5.1.2 得到了 $E(\varepsilon_i)=0$ 对所有的 i 都成立，当然有 $E(\varepsilon_i)=0$，$E(\varepsilon_j)=0$，因此，$\text{cov}(\varepsilon_i,\varepsilon_j)=E[\varepsilon_i-E(\varepsilon_i)][\varepsilon_j-E(\varepsilon_j)]=E(\varepsilon_i,\varepsilon_j)$。

根据假设，自然有 $\text{cov}(\varepsilon_i,\varepsilon_j)=E(\varepsilon_i,\varepsilon_j)=0$

如果 $E(\varepsilon_i,\varepsilon_j)\neq 0$，叫**自相关**（Autocorrelation）。当然，$E(\varepsilon_i,\varepsilon_j)=0$ 时，$E(\varepsilon_{i-1},\varepsilon_i)=0$ 和 $E(\varepsilon_i,\varepsilon_{i+1})=0$ 显然都成立。

假设 5.1.5　所有 x_i 都是可观察的并且独立于 ε_i 的，即对所有 i,j 来说，$\text{cov}(x_i,\varepsilon_j)=0$。

这就保证了 ε_i 值与 x_i 的取值无任何关系，请看下面这个例子：

ε_3 是对应 $x_{13},x_{23},\cdots,x_{k3}$ 的残差，即：

$$y_3=\alpha+\beta_1x_{13}+\beta_2x_{23}+\cdots+\beta_kx_{k3}+\varepsilon_3$$

上式中 $x_{13},x_{23},\cdots,x_{k3}$ 的取值不仅与 ε_3 毫无关系，也与其它 x_{ij} 无关。同时 ε_3 的值与 x_{i3} 无关，与其它 x_{ij} 也没有关系，即 ε_i 既与 x_{ij} 的变化无关，也不影响 x_{ij}。但是现实中这条假设是否满足可要打个大大的折扣了。

假设 5.1.6　数据产生过程是线性的，即：

$$y_i=\alpha+\beta_1x_{1i}+\beta_2x_{2i}+\cdots+\beta_kx_{ki}+\varepsilon_i$$

在满足以上古典假设的条件下，我们有必要进一步对因变量 y 的统计性质进行讨论。同时也有必要对模型中未知量 α 和 ε_i 最小二乘估计量的性质进行研究。这些性质之所以十分重要，是因为它使我们能够对一个建立在样本的观测数据基础上的 x 对 y 的影响关系进行检验。因此探讨这些最小二乘估计量的性质就十分必要了。

我们首先根据古典误差假设，推断出因变量的性质，再通过高斯－马尔科夫定理精确地讨论最小二乘估计量的一些性质。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第二节 古典假设的一些内涵

一、y_i 的分布

我们已知 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$，这里每一个 ε_i 都服从同样的独立分布，即 ε_i 对于 $i = 1,2,\cdots,n$ 的每一个观察值，方差为 σ^2，期望为0，也就是 $\varepsilon_i \sim i.i.d(0,\sigma^2)$，这里 i. i. d 指**独立同分布**（Identical Independent Distribution），这是古典假设保证成立的，但我们目前还不知道具体是哪一类分布。

正是由于 ε_i 的作用，引起 y_i 自身变化。求期望，有

$$E(y_i) = E(\alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i)$$

由于 $\alpha,\beta_1,\beta_2,\cdots,\beta_k$ 及 x_{ij} 都是常量，那么以 U_i 为常量，令

$$U_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}（注意：这里 U_i 不是随机变量）$$

根据假设5.1.2，$E(\varepsilon_i) = 0$，可知

$$E(y_i) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} = U_i$$

对于不同的 i，有不同的 $E(y_i)$。这说明 y_i 的期望是不断变化的，不是同分布的（实际上是一条直线）。

对于 y_i 的方差来说，$Var(y_i) = Var(\alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i)$（因为 $\alpha,\beta_1,\beta_2,\cdots,\beta_k$ 和 $x_{1i},\cdots,x_{ki}$ 都是常量）$= Var(\varepsilon_i) = \sigma^2$，也就是说，$y_i$ 与 ε_i 同方差，y_i 的变动幅度是与总体残差 ε_i 的变动幅度是一样的，如图5.5所示。

$$E(y_i) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}$$

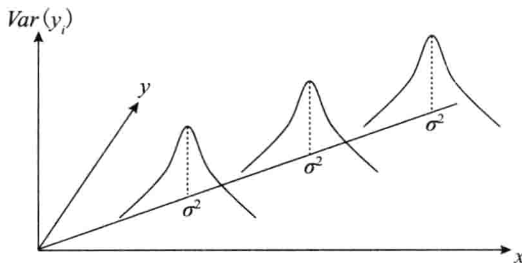


图5.5

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

我们应注意以下问题，由于这个方差不依赖于 $\alpha+\beta_1x_{1i}+\beta_2x_{2i}+\cdots+\beta_kx_{ki}$，其结果是它也不依赖于任何 x_{ij} 的任何样本值。同时，由于这是一个常量加到一个随机变量上，因此并不影响它与其它随机变量的相关程度，$\mathrm{cov}(y_i,y_j)$ 仍为零，就如 $\mathrm{cov}(\varepsilon_i,\varepsilon_j)=0(i\neq j)$ 一样。因此，在古典假设的条件下，由于 x_{ij} 的值是可知且独立于 ε_i 的，每一个 y_i 的分布在已知 x_{ij} 的情况下可以表示为：$y_i\sim i.i.d(\alpha+\beta_1x_{1i}+\beta_2x_{2i}+\cdots+\beta_kx_{ki},\sigma^2)$。

二、高斯-马尔科夫定理：最小二乘估计量的样本分布

利用给定的一组样本值，我们可以回归出一条力图反映母体中数据产生过程的 $y_i=\alpha+\beta_1x_{1i}+\beta_2x_{2i}+\cdots+\beta_kx_{ki}+\varepsilon_i$ 的直线，得到估计量 $\hat{\alpha}$、$\hat{\beta}_i$、$\hat{y}_i$。给另一组样本值，又可以回归出一条新的直线，得到不同的估计量 $\hat{\alpha}$、$\hat{\beta}_i$、$\hat{y}_i$。所给的样本组数越多，就能得到越多不同的估计量。但是，必须记住，我们所得出的每一条回归直线，必定多多少少反映了母体产生数据过程的性质特点。这就像一个家庭中，每个子女都多多少少像他们的父母一样。有的可能极像，有的稍差一些。正因为有这样生成数据过程的作用，决定了回归出的 $\hat{\alpha}$、$\hat{\beta}_i$、$\hat{y}_i$ 必然有一个变化的范围。我们绝不可能在一组样本中回归得到母体 α、β_i、y_i 的真实值。这也就是说，在 $\varepsilon_i\sim i.i.d(0,\sigma^2)$ 下，$\hat{\alpha}$、$\hat{\beta}$ 是随机分布的，但有自己的分布规律。

以 $y_i=\alpha+\beta x_i+\varepsilon_i$ 为例，由于

$$\hat{\beta}=\sum_{i=1}^{n}w_iy_i,\ \hat{\alpha}=\frac{1}{n}\sum y_i-\left(\sum w_iy_i\right)\bar{x}=\sum\left(\frac{1}{n}-\bar{x}w_i\right)y_i=\sum w_i^*y_i$$

说明了 $\hat{\alpha}$、$\hat{\beta}$ 也是随机变量。因为它们都是随机变量 y_i 的线性组合。作为随机变量，$\hat{\alpha}$、$\hat{\beta}_i$ 必定有各自的样本分布。这是我们下面将着手解决的问题。

（一）第一种类型：针对 $y_i=\alpha+\beta x_i+\varepsilon_i$

在 $\hat{\beta}=\dfrac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$ 中，先讨论它的分子。对分子进行变形可知

$$\begin{aligned}\sum(x_i-\bar{x})(y_i-\bar{y})&=\sum[(x_i-\bar{x})y_i-(x_i-\bar{x})\bar{y})]\\&=\sum(x_i-\bar{x})y_i-\bar{y}\sum(x_i-\bar{x})\\&=\sum(x_i-\bar{x})y_i-\bar{y}\left(\sum x_i-n\bar{x}\right)\end{aligned}$$

由于 $\bar{x}=\dfrac{\sum x_i}{n}$，则 $\sum x_i-n\cdot\bar{x}=0$，因此，$\sum(x_i-\bar{x})y_i-\bar{y}(\sum x_i-n\bar{x})=\sum(x_i$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$-\bar{x})y_i$，这是一种很常用的技巧。这时再来整体看 $\hat{\beta}$，它等于 $\dfrac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}$，其中 $\sum(x_i-\bar{x})^2$ 是可知量，所以 $\hat{\beta}$ 不妨写成：$\hat{\beta}=\sum\dfrac{(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}y_i$。

令 $w_i=\dfrac{(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$，则 $\hat{\beta}$ 变成了：$\hat{\beta}=\sum w_iy_i$，这说明 $\hat{\beta}$ 是线性的，相对于 y_i 而言。

仔细研究一下 w_i，可以知道它有如下性质：

（1）$\sum w_i=0$；（2）$\sum w_i^2=\dfrac{1}{\sum(x_i-\bar{x})^2}$；（3）$\sum w_i(x_i-\bar{x})=1$；（4）$\sum w_ix_i=1$。

证明：（1）论证 $\sum w_i=0$ 如下：

$$\begin{aligned}\sum w_i&=\sum\frac{x_i-\bar{x}}{\sum(x_i-\bar{x})^2}\\&=\frac{\sum(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\\&=\frac{\sum x_i-n\bar{x}}{\sum(x_i-\bar{x})^2}\\&=\frac{n(\frac{\sum x_i}{n}-\bar{x})}{\sum(x_i-\bar{x})^2}\\&=\frac{n(\bar{x}-\bar{x})}{\sum(x_i-\bar{x})^2}\\&=0\end{aligned}$$

（2）论证 $\sum w_i^2=\dfrac{1}{\sum(x_i-\bar{x})^2}$ 如下：

$$\begin{aligned}\sum w_i^2&=\sum\left[\frac{x_i-\bar{x}}{\sum(x_i-\bar{x})^2}\right]^2\\&=\frac{\sum(x_i-\bar{x})^2}{\left[\sum(x_i-\bar{x})^2\right]^2}\end{aligned}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$= \frac{1}{\sum (x_i - \bar{x})^2}$$

（3）论证 $\sum w_i(x_i - \bar{x}) = 1$ 如下：

$$\sum w_i(x_i - \bar{x}) = \sum \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}(x_i - \bar{x})$$

$$= \sum \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}$$

$$= \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}$$

$$= 1$$

（4）论证 $\sum w_i x_i = 1$ 如下：

因为 $\sum w_i(x_i - \bar{x}) = \sum w_i x_i - \bar{x}\sum w_i$，而 $\sum w_i = 0$。这是我们论证过了的结论（1）。也就是说，$\sum w_i x_i$ 肯定等于 $\sum w_i(x_i - \bar{x})$，而 $\sum w_i(x_i - \bar{x}) = 1$，那么 $\sum w_i x_i = 1$ 也就不言而喻了。

下面由 w_i 的四个性质，求出一元回归模型后计算 $\hat{\alpha}$、$\hat{\beta}$ 的方差与期望：

1. 方差 $E(\hat{\beta})$

$E(\hat{\beta}) = E(\sum w_i y_i) = \sum w_i[E(y_i)]$。因为 w_i 为一常量，可以提到 E 号外，而我们在本章第二节中首先讨论了 y_i 的分布。针对一元方程 $y_i = \alpha + \beta x_i + \varepsilon_i$，$E(y_i) = \alpha + \beta x_i$。代入到 $E(\hat{\beta}) = \sum w_i[E(y_i)]$ 中去，可知：

$$E(\hat{\beta}) = \sum w_i[E(y_i)]$$

$$= \sum w_i(\alpha + \beta x_i)$$

$$= \sum (w_i\alpha + w_i\beta x_i)$$

$$= \alpha\sum w_i + \beta\sum x_i w_i$$

由 w_i 的性质（1）$\sum w_i = 0$ 和性质（4）$\sum w_i x_i = 1$，得到：$E(\hat{\beta}) = \beta$。

这说明最小二乘估计量 $\hat{\beta}$ 是 β 的无偏估计量。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

2. 期望 $Var(\hat{\beta})$

$$
\begin{aligned}
Var(\hat{\beta}) &= Var(\sum w_i y_i) \\
&= \sum w_i^2 Var(y_i) \text{（由方差中提累加值必须变成平方累加值）} \\
&= \sum w_i^2 \sigma^2 \\
&= \sigma^2 \sum w_i^2
\end{aligned}
$$

由 w_i 的性质（2）可知：$\sum w_i^2 = \frac{1}{\sum (x_i - \bar{x})^2}$，因此

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

这样我们就得到了估计量 $\hat{\beta}$ 的方差与期望。同理可以得到估计量 $\hat{\alpha}$ 的方差和期望，方法类似。这里仅仅给出结果，不再列示求解步骤了。

$$E(\hat{\alpha}) = \alpha,\ Var(\hat{\alpha}) = \left[\frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} + \frac{1}{n}\right]\sigma^2$$

这样，对模型 $y_i = \alpha + \beta x_i + \varepsilon_i$，我们能得到如下结论：

(1) α 的性质：$E(\hat{\alpha}) = \alpha,\ Var(\hat{\alpha}) = \left[\frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} + \frac{1}{n}\right]\sigma^2$；

(2) β 的性质：$E(\hat{\beta}) = \beta,\ Var(\hat{\beta}) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$；

(3) α、β 的关系：$\text{cov}(\hat{\alpha},\hat{\beta}) = -\frac{\bar{x}}{\sum (x_i - \bar{x})^2}\sigma^2$。

（二）第二种类型：针对 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$。

求导二元回归的方法较繁，这里仅给出结果：

(1) $\hat{\alpha}$ 的性质：$E(\hat{\alpha}) = \alpha$；

(2) $\hat{\beta}_1$ 的性质：$E(\hat{\beta}_1) = \beta_1,\ Var(\hat{\beta}_1) = \sigma^2 \frac{\sum \dot{x}_{2i}^2}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$；

(3) $\hat{\beta}_2$ 的性质：$E(\hat{\beta}_2) = \beta_2,\ Var(\hat{\beta}_2) = \sigma^2 \frac{\sum \dot{x}_{1i}^2}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第三节　高斯-马尔科夫定理

一、高斯-马尔科夫定理（The Guass - Markov Theorem）基本概念

高斯-马尔科夫定理：对模型 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$，$\varepsilon_i \sim i.i.d(0,\sigma^2)$，最小二乘估计量 $\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2,\cdots,\hat{\beta}_k$ 是 $\alpha,\beta_1,\beta_2,\cdots,\beta_k$ 最好的、线性的、无偏的估计量。

所谓最好的估计量，是指估计量中方差最小的那个估计量。也就是说，最小二乘法得到的估计量在所有线性无偏估计量中，方差最小。

所谓线性的估计量，是指估计量与 y_i 存在线性关系。

所谓无偏的估计量，是指估计量的期望值等于真实值。

以 $\hat{\beta}_i$ 为例，所谓最好的是指方差最小，即 $\mathrm{Min}Var(\hat{\beta}_i)$；所谓无偏的是指 $E(\hat{\beta}_i) = \beta_i$；所谓线性的是指 $\hat{\beta}_i = \sum C_i y_i$。

二、以模型 $y_i = \alpha + \beta x_i + \varepsilon_i$，$\varepsilon_i \sim i.i.d(0,\sigma^2)$ 为例证明高斯-马尔科夫定理

实际上，我们在推导 $\hat{\beta}$、$\hat{\alpha}$ 的期望和方差时，已经得出了结论，现以 $\hat{\beta}$ 为例。

证明：（1）线性

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$= \sum \frac{(x_i - \bar{x})y_i}{\sum (x_i - \bar{x})^2}$$

$$= \sum w_i y_i，其中\ w_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}。$$

因此，$\hat{\beta}$ 为 y_i 的线性估计量。

（2）无偏

$$E(\hat{\beta}) = E(\sum w_i y_i)$$

$$= \sum w_i E(y_i)$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$= \sum w_i(\alpha + \beta x_i)$$

$$= \alpha \sum w_i + \beta \sum w_i x_i$$

由 w_i 性质（1）、（4），$\sum w_i = 0$，$\sum w_i x_i = 1$，得：$E(\hat{\beta}) = \beta$

（3）最好

思路1： 欲证 $\hat{\beta}$ 最好，即要证明 $Var(\hat{\beta}_i)$ 是所有线性估计量中方差最小的，也就是说设法证明其它线性估计量的方差都大于 $\hat{\beta}$ 的方差。不妨取估计量 $\breve{\beta}$，令 $\breve{\beta} = \sum (w_i + B_i)y_i$，当 $B_i = 0, i = 1,2,\cdots,n$ 时，$\hat{\beta} = \breve{\beta}$。无论是 $\breve{\beta}$ 还是 $\hat{\beta}$，都是在无偏与线性约束条件下有关 β 的估计量，所以 $\breve{\beta}$ 必须是线性、无偏的。因此我们从中可得一些 B_i 的特点，从而使证明 $\hat{\beta}$ 为最好转为证明 $Var(\breve{\beta}) \geq Var(\hat{\beta})$）。

B_i 有什么特点呢？

$$\begin{aligned} E(\breve{\beta}) &= E[\sum (w_i + B_i)y_i] \\ &= \sum (w_i + B_i)[E(y_i)] \\ &= \sum (w_i + B_i)(\alpha + \beta x_i) \\ &= \alpha \sum w_i + \beta \sum w_i x_i + \alpha \sum B_i + \beta \sum B_i x_i \end{aligned}$$

由于 $\sum w_i = 0$，$\sum w_i x_i = 1$，所以 $E(\breve{\beta}) = \beta + \alpha \sum B_i + \beta \sum B_i x_i$。

为使 $E(\breve{\beta}) = \beta$，保证无偏，有：$\alpha \sum B_i + \beta \sum B_i x_i = 0$。

对于任意的 α、β，为使 $E(\breve{\beta})$，只有当 $\begin{cases} \alpha \sum B_i = 0 \\ \beta \sum B_i x_i = 0 \end{cases}$ 成立时，才能保证 α、β 之间的任意性。因此，从无偏性我们可以得知有 $\begin{cases} \sum B_i = 0 \\ \sum B_i x_i = 0 \end{cases}$ 条件成立。

下面就证明 $Var(\breve{\beta}) \geq Var(\hat{\beta})$：

$$\begin{aligned} Var(\breve{\beta}) &= Var[\sum (w_i + B_i)y_i] \\ &= \sum (w_i + B_i)^2 Var(y_i) \\ &= \sum (w_i^2 + 2w_iB_i + B_i^2) \cdot \sigma^2 \end{aligned}$$（由于 y_i 与 ε_i 同方差，才有 $Var(y_i) = \sigma^2$）

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$= \sum (w_i^2 + B_i^2)\sigma^2 + 2\sigma^2 \sum w_i B_i$$

现在，到了这一步后，可以看到：如果能够有 $Var(\breve{\beta}) = \sum (w_i^2 + B_i^2)\sigma^2 \geqslant \sum w_i^2\sigma^2$，而 $Var(\hat{\beta}) = \sum w_i^2\sigma^2$，那么一切都会迎刃而解。所以现在的问题是设法证明 $2\sigma^2 \sum w_i B_i = 0$ 即可。

$$\begin{aligned}\sum w_i\beta_i &= \sum \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2} B_i \\ &= \frac{\sum x_i B_i - \bar{x}\sum B_i}{\sum (x_i - \bar{x})^2} \\ &= \frac{0 - \bar{x}\cdot 0}{\sum (x_i - \bar{x})^2} \\ &= 0\end{aligned}$$

因此，
$$\begin{aligned}Var(\breve{\beta}) &= \sum (w_i^2 + B_i^2)\sigma^2 + 2\sigma^2 \sum w_i B_i \\ &= \sum (w_i^2 + B_i^2)\sigma^2 + 2\sigma^2 \cdot 0 \\ &= \sum (w_i^2 + B_i^2)\sigma^2 \geqslant \sum w_i^2\sigma^2 = Var(\hat{\beta})\end{aligned}$$

即：$Var(\breve{\beta}) \geqslant Var(\hat{\beta})$，所以 $\hat{\beta}$ 最好。

思路 2：由 $\breve{\beta}$ 为线性无偏约束可知：

$$\begin{aligned}E(\breve{\beta}) &= \sum B_i[E(y_i)] \\ &= \sum B_i(\alpha + \beta x_i) \\ &= \alpha\sum B_i + \beta\sum B_i x_i\end{aligned}$$

为保证无偏性，$E(\breve{\beta}) = \beta$，必有 $\begin{cases}\sum B_i = 0 \\ \sum B_i x_i = 0\end{cases}$

$$\begin{aligned}Var(\breve{\beta}) &= \sum B_i^2[Var(y_i)] \\ &= \sum B_i^2\sigma^2 \\ &= \sigma^2\sum B_i^2\end{aligned}$$

我们可以使用求条件极值的方法使这一方差最小。这样，同样能够证明这一最小方

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

差，即为最小二乘估计量的方差。

欲使 $\sigma^2 \sum B_i^2$ 最小，因 σ^2 已为定值，只须使得 $\sum B_i^2$ 最小即可。同时，若能够再证明 $\sum B_i = \sum w_i$，就证明了 $\breve{\beta} = \sum B_i y_i = \hat{\beta}$。

这样整个问题就转化为求下面的条件极值：

$$\underset{\beta_i}{\text{Min}} \sum B_i^2$$

$$\text{ST.} \sum B_i = 0$$

$$\sum B_i x_i = 1$$

对这个条件极值，我们可用拉格朗日方法来求解。具体步骤如下：

首先，设立拉格朗日函数为：$L(B_i, \lambda_1, \lambda_2) = \sum B_i^2 - \lambda_1 \sum B_i - \lambda_2 (\sum B_i x_i - 1)$。

对该函数求导，可知：

$$\begin{cases} \dfrac{\partial L}{\partial B_1} = 2B_1 - \lambda_1 B_1 - \lambda_2 x_1 = 0 & (1) \\ \dfrac{\partial L}{\partial B_2} = 2B_2 - \lambda_1 B_2 - \lambda_2 x_2 = 0 & (2) \\ \vdots & \\ \dfrac{\partial L}{\partial B_n} = 2B_n - \lambda_1 B_n - \lambda_2 x_n = 0 & (n) \\ \dfrac{\partial L}{\partial \lambda_1} = \sum B_i = 0 & (n+1) \\ \dfrac{\partial L}{\partial \lambda_2} = \sum B_i x_i - 1 = 0 & (n+2) \end{cases}$$

其次，导出 λ_1, λ_2 的关系：

$(1) + (2) + \cdots + (n)$ 可得，$2\sum B_i - n\lambda_1 - \lambda_2 \sum x_i = 0$。

由于已知 $\sum B_i = 0$，所以 $\lambda_1 = -\lambda_2 \dfrac{\sum x_i}{n} = -\lambda_2 \bar{x}$。

再次，导出 λ_2：

将 $\lambda_1 = -\lambda_2 \bar{x}$ 代回原式中：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\begin{cases} 2B_1 + \lambda_2\bar{x} - \lambda_2 x_1 = 0 \\ 2B_2 + \lambda_2\bar{x} - \lambda_2 x_2 = 0 \\ \vdots \\ 2B_n + \lambda_2\bar{x} - \lambda_2 x_n = 0 \end{cases}$$

移项后，可得

$$\begin{cases} 2B_1 = \lambda_2(x_1 - \bar{x}) \\ 2B_2 = \lambda_2(x_2 - \bar{x}) \\ \vdots \\ 2B_n = \lambda_2(x_n - \bar{x}) \end{cases}$$

则上式通项为 $B_i = \dfrac{\lambda_2}{2}(x_i - \bar{x})$ 。

在通项两边同乘以 x_i，可得 $x_iB_i = \dfrac{\lambda_2}{2}(x_i - \bar{x})x_i$ ；

对上式求累和，可得 $2\sum x_iB_i = \lambda_2\sum(x_i - \bar{x})x_i$ ；

运用约束条件 $\sum B_ix_i = 1$，则 $\lambda_2\sum(x_i - \bar{x})x_i = 2$，则 $\lambda_2 = \dfrac{2}{\sum(x_i - \bar{x})x_i}$ 。

最后，求出 $\breve{\beta}$ 的具体形式：

通项两边同乘 y_i，可得 $y_iB_i = \dfrac{\lambda_2}{2}(x_i - \bar{x})y_i$ ；

对上式求和，可得 $\sum B_iy_i = \dfrac{\lambda_2}{2}\sum(x_i - \bar{x})y_i$ ；

由于 $\breve{\beta} = \sum B_iy_i$，则

$$\breve{\beta} = \sum B_iy_i = \frac{\lambda_2}{2}\sum(x_i - \bar{x})y_i = \frac{1}{2}\cdot\frac{2}{\sum(x_i - \bar{x})x_i}\sum(x_i - \bar{x})y_i$$

（代入 $\lambda_2 = \dfrac{2}{\sum(x_i - \bar{x})x_i}$）

这里再作一个小技巧变形：

$$\sum(x_i - \bar{x})^2 = \sum(x_i - \bar{x})(x_i - \bar{x})$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$= \sum[(x_i - \bar{x})x_i - (x_i - \bar{x})\bar{x}]$$
$$= \sum(x_i - \bar{x})x_i - \bar{x}\sum(x_i - \bar{x})$$
$$= \sum(x_i - \bar{x})x_i - \bar{x}(\sum x_i - n\bar{x})$$
$$= \sum(x_i - \bar{x})x_i$$

这样，$\lambda_2 = \dfrac{2}{\sum(x_i - \bar{x})x_i} = \dfrac{2}{\sum(x_i - \bar{x})^2}$；

因此，$\breve{\beta} = \dfrac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}$。

而 $\sum(x_i - \bar{x})y_i$ 又令我们想起什么呢？

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - \bar{x})y_i - \bar{y}\sum(x_i - \bar{x})$$
$$= \sum(x_i - \bar{x})y_i - \bar{y}(\sum x_i - n\bar{x})$$
$$= \sum(x_i - \bar{x})y_i - \bar{y}\cdot 0$$
$$= \sum(x_i - \bar{x})y_i$$

上面的这个技巧反反复复地使用，读者应该很熟悉了。这样

$$\breve{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \hat{\beta}$$

这意味着，我们所得到的方差最小的估计量，即所有线性无偏估计量中最好的实际上就是使用最小二乘法所得到的估计量。

第四节　高斯-马尔科夫定理再探

在现实世界中，探索种种经济规律和变化趋势时，我们一般总是通过一些实验来发现各种差异，进而对差异进行分析和研究，找出经济规律在不同情况下的作用特征，然后将这些不同特征进行回归分析后，再找出它们的内在规律。在我们所进行的实验中，有些是可控制的实验，从而得到的数据也是可控制的数据，有些是不可以控制的实验，其数据自然也是不可控制的数据。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

可控实验中，你把自变量取值的差距拉得越大越好，也就是说，自变量的方差越大，越能说明问题。我们上面所做的一切，实际上，都是在对参数估计量性质作说明。

在第四章中引进 V_{1i} 时，我们定义 $V_{1i}=x_{1i}-\hat{x}_{1i}$，它是 x_{1i} 对变量 x_{2i} 回归后的残差，现在我们又由此导出二元回归方程中估计量 $\hat{\beta}_1$ 的方差的另一种形式：

$$Var(\hat{\beta}_1)=\frac{\sigma^2}{(1-R^2)\sum(x_{1i}-\bar{x}_1)^2}$$

现在把它推广：由原先的 $x_{1i}=\alpha+\beta x_{2i}$，推广到现在的 $x_{ij}=\alpha+\beta_1x_{1j}+\cdots+\beta_{i-1}x_{i-1j}+\beta_{i+1}x_{i+1j}+\cdots+\beta_kx_{kj}$，仍然定义 $V_{ij}=x_{ij}-\hat{x}_{ij}$ 这里 V_{ij} 表示的是 x_i 对其它所有自变量 $x_1,x_2,\cdots,x_{i-1},x_{i+1},\cdots,x_k$ 进行回归得到的拟合残差。那么，多元线性回归模型 $y_i=\alpha+\beta_1x_{1i}+\cdots+\beta_kx_{ki}$ 中估计量 $\hat{\beta}_i$ 的方差为：

$$Var(\hat{\beta}_i)=\frac{\sigma^2}{(1-R_i)^2\sum(x_{ij}-\bar{x}_i)^2}$$

其中 R_i^2 是 x_i 对 $x_1,x_2,\cdots,x_{i-1},x_{i+1},\cdots,x_k$ 回归时得到的相关系数。$\sum(x_{ij}-\bar{x}_i)^2$ 是上述回归中的 TSS。

上面的公式更具有普遍性，也好记多了。但是，我们仍可看出，公式中分母上的两个减项会给我们带来许多麻烦。

在以下这两种情况下会得不到 $Var(\hat{\beta}_1)$：

1. 离差为零，即 $\sum(x_{ij}-\bar{x}_i)^2=0$；
2. 多重共线性：即 $R_i^2=1$。

下面我们就来解决它们。

一、自变量离差为零（$\sum(x_{ij}-\bar{x}_i)^2=0$）

这个问题的解决办法相对来说较容易。对可控制实验而言，既然实验是可控的。我们可以在设计实验时尽量把 x_i 的方差拉大一些就可以了。对不可控实验而言，我们在搜集数据时，应尽量把 x_i 的方差拉大。

二、多重共线性

这个问题在处理时就相对复杂多了。在实验中我们将面对两种情况：

1. 我们确切地知道 R_i^2 就等于 1；

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

2. 我们知道 R_i^2 接近于 1（$R_i^2 \doteq 1$）。

删去变量，是处理多重共线性的好方法。当第一种情况 $R_i^2=1$ 存在时，应该删去这个变量。但在第二种情况，即 $R_i^2 \doteq 1$ 时，问题来了。因为对 $R_i^2=1$ 不确定，可能产生下面两个麻烦：

（1）在生成过程中原本应包括这个变量，但由于我们认为 $R_i^2=1$ 成立而给删去了，这叫变量的误删（Underspecification of a Regression Equation）。

（2）在生成过程中原本不包括这个变量，但由于我们认为 $R_i^2 \neq 1$ 成立，导致该变量的误加（Overspecification）。

上面的这两个问题，可以用表 5－1 表示：

表 5－1

后果 / 变量在数据生产过程中 / 我们做出的决定	包括（Included）	不包括（Excluded）
删去（Drop）	变量的误删	决定正确
保留（Keep）	决定正确	变量的误加

下面，我们专门讨论一下有关变量的误删与误加问题。

（一）变量的误删（Underspecification of a Regression Equation）

假如真正的数据生成过程为：

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \varepsilon_i \sim i.i.d(0,\sigma^2)$$

由于种种原因，如：数据掌握不足，使用的定理范围限制等，我们觉得生成过程为：

$$y_i = a + b_1 x_{1i} + \varepsilon_i, \varepsilon_i \sim i.i.d(0,\sigma^2)$$

现在研究 $\hat{b}_1$ 的均值和方差：

$$\hat{b}_1 = \frac{\sum (x_{1i}-\bar{x}_1)(y_i-\bar{y})}{\sum (x_{1i}-\bar{x}_1)^2} = \frac{\sum [(x_{1i}-\bar{x}_1)y_i - \bar{y}(x_{1i}-\bar{x}_1)]}{\sum (x_{1i}-\bar{x}_1)^2}$$

$$= \frac{\sum (x_{1i}-\bar{x}_1)y_i - \bar{y}\sum (x_{1i}-\bar{x}_1)}{\sum (x_{1i}-\bar{x}_1)^2} = \frac{\sum (x_{1i}-\bar{x}_1)y_i - \bar{y}(\sum x_{1i} - n\bar{x}_1)}{\sum (x_{1i}-\bar{x}_1)^2}$$

$$= \frac{\sum (x_{1i}-\bar{x}_1)y_i}{\sum (x_{1i}-\bar{x}_1)^2}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

首先考察 $\hat{b}_1$ 的期望：

$$E(\hat{b}_1)=\frac{\sum(x_{1i}-\bar{x}_1)E(y_i)}{\sum(x_{1i}-\bar{x}_1)^2}$$

$$=\frac{\sum(x_{1i}-\bar{x}_1)(\alpha+\beta_1x_{1i}+\beta_2x_{2i})}{\sum(x_{1i}-\bar{x}_1)^2}$$

$$=\frac{\alpha\sum(x_{1i}-\bar{x}_1)+\beta_1\sum(x_{1i}-\bar{x}_1)x_{1i}+\beta_2\sum(x_{1i}-\bar{x}_1)x_{2i}}{\sum(x_{1i}-\bar{x}_1)^2}$$

由于 $\sum(x_i-\bar{x})=\sum x_i-n\bar{x}=n(\frac{\sum x_i}{n}-\bar{x})=n(\bar{x}-\bar{x})=0$，

则：$E(\hat{b}_1)=\beta_1+\beta_2\dfrac{\sum(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{\sum(x_{1i}-\bar{x}_1)^2}$。

这说明 $\hat{b}_1$ 是 β_1 的一个有偏估计量，造成有偏的原因是因为丢失了自变量（这里是指 x_{2i}）。该偏差为：$Bias(\hat{b}_1)=\beta_2\dfrac{\sum(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{\sum(x_{1i}-\bar{x}_1)^2}$。

需要对以上推导加以说明的是：

由于是实际的 y_i，所以代入 $E(y_i)=\alpha+\beta_1x_{1i}+\beta_2x_{2i}$，而不是代入我们假想的 $y_i=a+b_2x_{1i}+\varepsilon_i$。再看方差：$Var(\hat{b}_1)=\dfrac{\sigma^2}{\sum(x_{1i}-\bar{x}_1)^2}$，而 $Var(\hat{\beta}_1)=\dfrac{\sigma^2}{(1-R_1^2)\sum(x_{1i}-\bar{x}_1)^2}$③。

显然，$Var(\hat{b}_1)\leq Var(\hat{\beta}_1)$。

综上所述，变量丢失会产生以下后果：

（1）估计量 $\hat{b}_1$ 有偏，即 $E(\hat{b}_1)\neq\beta_1$；

（2）估计量 $\hat{b}_1$ 方差减小：$Var(\hat{b}_1)\leq Var(\hat{\beta}_1)$。

这种情况用图 5.6 表示就是：

③ 由于省略多个变量的变量丢失问题不可避免要用到矩阵法，这里就不列出了，其证明可在许多计量经济学教科书中找到，例如：G. S. Maddala：Econometrics，New York：McGraw—Hill，1977，P. 461.

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



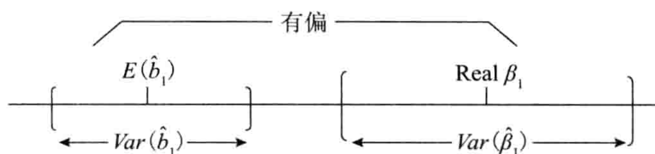


图 5.6

（二）变量的误加（Overspecification of a Regression Equation）

在实际中，我们经常考虑在回归模型中加入每个可能的自变量，以此希望能增强模型的解释能力（R^2），或改进预测准确度。尽管这类模型有时能提供短期的确实可信的预报，但这种做法也极容易产生多余变量的误加问题。下面，我们就对这一问题专门讨论一下。

真正的数据生成过程为：

$$y_i = \alpha + \beta x_{1i} + \varepsilon_i, \varepsilon_i \sim i.i.d(0,\sigma^2)$$

估计时使用的生成过程为：

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + \varepsilon_i, \varepsilon_i \sim i.i.d(0,\sigma^2)$$

同讨论变量丢失问题一样，我们先求出 $E(\hat{b}_1)$ 和 $Var(\hat{b}_1)$：

$$\hat{b}_1 = \frac{\sum \dot{x}_{2i}^2 \sum \dot{x}_{1i}\dot{y}_i - \sum \dot{x}_{1i}\dot{x}_{2i} \sum \dot{x}_{2i}\dot{y}_i}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$$

由于 $\sum \dot{x}_{1i}\dot{y}_i = \sum (x_{1i} - \bar{x}_1)(y_i - \bar{y})$

$$= \sum (x_{1i} - \bar{x}_1) y_i - \bar{y} \sum (x_{1i} - \bar{x}_1)$$

$$= \sum (x_{1i} - \bar{x}_1) y_i$$

则 $E(\sum \dot{x}_{1i}\dot{y}_i) = E[\sum (x_{1i} - \bar{x}_1) y_i]$

$$= \sum (x_{1i} - \bar{x}_1) E(y_i)$$

$$= \sum (x_{1i} - \bar{x}_1)(\alpha + \beta x_{1i})$$

$$= \alpha \sum (x_{1i} - \bar{x}_1) + \beta \sum (x_{1i} - \bar{x}_1) x_{1i}$$

$$= \beta \sum (x_{1i} - \bar{x}_1)^2$$

$$= \beta \sum \dot{x}_{1i}^2$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

同理 $E(\sum \dot{x}_{2i}\dot{y}_i) = \sum (x_{2i} - \bar{x}_2)(\alpha + \beta x_{1i})$

$$= \beta \sum (x_{2i} - \bar{x}_2) x_{1i}$$

$$= \beta \sum (x_{2i} - \bar{x}_2)(x_{1i} - \bar{x}_1)$$

$$= \beta \sum \dot{x}_{1i}\dot{x}_{2i}$$

因此 $E(\hat{b}_1) = E\dfrac{\sum \dot{x}_{2i}^2 \sum \dot{x}_{1i}\dot{y}_i - \sum \dot{x}_{1i}\dot{x}_{2i} \sum \dot{x}_{2i}\dot{y}_i}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$

$$= \frac{\sum \dot{x}_{2i}^2 E(\sum \dot{x}_{1i}\dot{y}_i) - \sum \dot{x}_{1i}\dot{x}_{2i} E(\sum \dot{x}_{2i}\dot{y}_i)}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$$

$$= \frac{\sum \dot{x}_{2i}^2 (\beta \sum \dot{x}_{1i}^2) - \sum \dot{x}_{1i}\dot{x}_{2i} (\beta \sum \dot{x}_{1i}\dot{x}_{2i})}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$$

$$= \beta$$

现在再看 $Var(\hat{b}_1)$ 和 $Var(\hat{\beta})$：

$$Var(\hat{b}_1) = \frac{\sigma^2}{(1 - R_1{}^2) \sum (x_{1i} - \bar{x}_1)^2}$$

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum (x_{1i} - \bar{x}_1)^2}$$

很明显，$Var(\hat{b}_1) \geq Var(\hat{\beta})$。

这样，我们可以得到变量误加结果如下：

（1）估计量 $\hat{b}_1$ 无偏：$E(\hat{b}_1) = \beta$；

（2）估计量 $\hat{b}_1$ 方差较大：$Var(\hat{b}_1) \geq Var(\hat{\beta})$。

图示见图 5.7 所示：

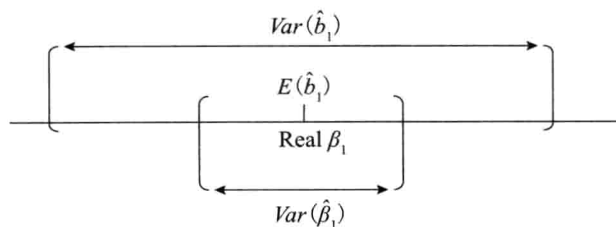


图 5.7

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



这样看来，人们会认为变量误加要比变量丢失后果好得多。但实际上，由于各种各样的其它原因，事情并非如此简单。

在经济生活中，经济数据总是不规则地反映经济规律的作用结果。但这些数据的产生不是完全任意的，在其背后有一个真正的数据生成过程。我们的任务是从母体的一组样本中，利用已知的离散点推测其真正的数据生成过程。这种推导有两种方法：点估计方法和区间估计方法。我们目前使用的方法是点估计方法中的最小二乘法。

因此，我们必须看到，以 β_1 为例，其估计值可能会有许多个。它可以通过任何估计方法得到，可能是线性估计量也可能是非线性估计量。但是，无论如何，这些估计量中必须有一些估计量为线性估计量。当然，在这些线性估计量中，有的是有偏的，有的是无偏的，如图 5.8 所示：

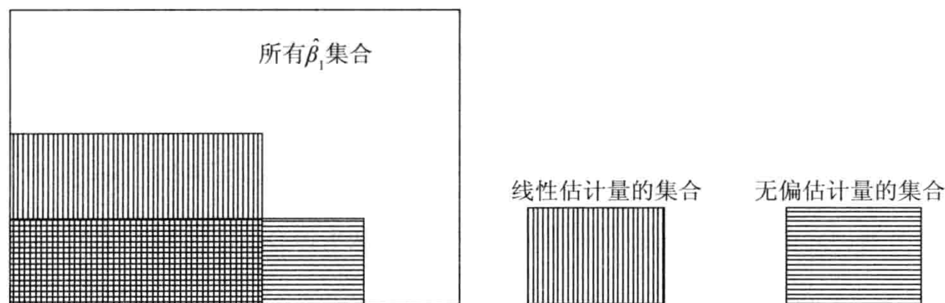


图 5.8

在重合部分，即线性一元无偏估计量中，有一个最好的估计量。这个最好的估计量被马尔科夫定理认为就是在最小二乘法中得到的估计量。

如何证明它是最好的呢？在前面，我们曾使用两种方法：第一种，我们认定最小二乘法下得到的 β 是最好的，可以在重合区间中，再找另外一个线性无偏估计量，证明其方差总比最小二乘估计量的方差大。这是顺证法，无非是 $\hat{\beta}_i$ 本身再加一点，减一点，就可以得到另外一个估计量 $\hat{\beta}_i = \sum (w_i + B_i) y_i$，然后再比较其方差。

第二种方法是逆证法。我们干脆撇开 $\hat{\beta}$，自己先找到最好的估计量，然后论证这个最好的估计量就是最小二乘法下所得到的估计量。此时使用了约束条件下求极值的方法。

目前我们已经掌握了上面这些知识，并得知了回归模型中各未知参数的方差。特别是，针对二元回归模型 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ 曾直接给出了 $\hat{\beta}_1$、$\hat{\beta}_2$ 的方差：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$Var(\hat{\beta}_1) = \sigma^2 \frac{\sum \dot{x}_{2i}^2}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$$

$$Var(\hat{\beta}_2) = \sigma^2 \frac{\sum \dot{x}_{1i}^2}{\sum \dot{x}_{1i}^2 \sum \dot{x}_{2i}^2 - (\sum \dot{x}_{1i}\dot{x}_{2i})^2}$$

在以后的学习中，这两个公式并不十分常用。而是使用另外一种变形，以 $\hat{\beta}_1$ 为例：

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{(1 - R_1^2) \sum (x_{1i} - \bar{x}_1)^2}$$

怎么推导出的呢?

首先回顾一下我们在前面专门讨论的消除残差法的内容（建议读者再翻回去看一看“消除了 x_{2i} 影响后的 v_{1i}，即 x_{1i} 自身变化对 y_i 的作用，可用新的一元回归方程来拟合”），也就是对 $y_i = f + gv_{1i} + \varepsilon_i$ 进行拟合。而我们后来已证明了 $\hat{g}$ 实际上就是二元回归方程中的 $\hat{\beta}$。就“式”论“表”，求一下 $\hat{g}$ 的方差。根据一元方程的参数公式，应用 $Var(\hat{g}) = \frac{\sigma^2}{\sum (v_{1i} - \bar{v}_{1i})^2}$。因为 $\bar{v}_{1i} = 0$，这是我们当时已证明的。所以：$Var(\hat{g}) = \frac{\sigma^2}{\sum v_{1i}^2}$。

触类旁通，之前提到 $v_{1i} = x_{1i} - \hat{x}_{1i}$，也就是真实值减去拟合值，实际上就是一元线性回归模型 $x_{1i} = c + dx_{2i} + \eta_i$ 的拟合残差。那么：$Var(\hat{g}) = \frac{\sigma^2}{\sum (x_{1i} - \bar{x}_{1i})^2} = \frac{\sigma^2}{ESS_1}$，其中 ESS_1 为一元回归模型 $x_{1i} = c + dx_{2i} + \eta_i$ 的 ESS，这是关键的一步。

还记得 ESS、TSS、R^2 的公式吧? $R_1^2 = 1 - \frac{ESS_1}{TSS_1}$。

因此，$ESS_1 = (1 - \mathrm{R}_1)^2 TSS_1$。代入上式，可知：$Var(\hat{g}) = \frac{\sigma^2}{ESS_1} = \frac{\sigma^2}{(1 - R_1^2) TSS_1}$，$TSS_1 = \sum (x_{1i} - \bar{x}_{1i})^2$。

因此，我们可以得到 $Var(\hat{\beta}_1) = Var(\hat{g}) = \frac{\sigma^2}{(1 - R_1^2) \sum (x_{1i} - \bar{x}_{1i})^2}$。

到目前为止，我们已经学会了许多东西，应该在实际中操作一下了。这一章里准备给读者一个自己解决二元回归问题的机会，可以先试试，再看看书上的计算步骤，如果过分依赖计算机，脑子里并不会得到充实。

假设数据生成过程为：$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \varepsilon_i \sim i.i.d(0, \sigma^2)$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



已知 y_i，x_{1i} 和 x_{2i} 的 10 个样本值，见表 5－2：

表 5－2

y_i	x_{1i}	x_{2i}
10. 939	6. 031	2. 041
9. 507	4. 973	1. 813
9. 078	7. 862	3. 882
7. 691	6. 669	3. 549
11. 753	4. 407	0. 687
8. 651	4. 416	1. 547
3. 795	9. 455	6. 705
12. 089	4. 161	0. 411
6. 573	5. 027	2. 827
8. 061	4. 137	1. 739

要求：（1）用二元线性回归模型 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$，得到 $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$，$Var(\hat{\beta}_1)$，$Var(\hat{\beta}_2)$，$ESS$，$RSS$，$R^2$；（2）用一元线性回归模型：$y_i = \alpha + \beta_1^* x_{1i} + \varepsilon_i$，得到 $\hat{\alpha}, \hat{\beta}_1^*$，$Var(\hat{\alpha})$，$Var(\hat{\beta}_1^*)$，$ESS$，$RSS$，$R^2$。

第四节实际主要讨论了线性回归模型中估计量的方差问题。

对于生成数据过程 $y_i = \alpha + \beta x_{1i} + \varepsilon_i$ 来说，$Var(\hat{\beta}) = \dfrac{\sigma^2}{\sum (x_{1i} - \bar{x}_1)^2}$；

对于生成数据过程 $y_i = \alpha + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i$ 来说，$Var(\hat{\beta}_i) = \dfrac{\sigma^2}{(1 - R_i^2) \sum (x_{ij} - \bar{x}_i)^2}$。

在求方差时，易产生的两个问题是：

（1）$\sum_j (x_{ij} - \bar{x}_i)^2 = 0$；（2）$R_i^2 = 1$。

图 5.9 有助于掌握它们之间的逻辑结构。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

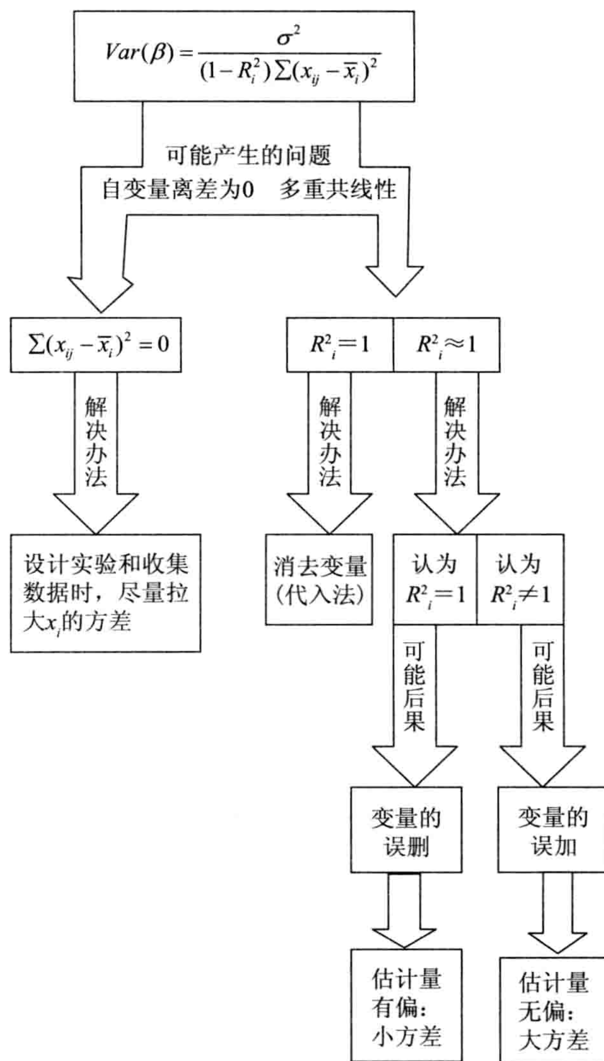


图 5.9

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第六章 正态条件下回归的推论

在以前讨论线性回归问题时，模型 $y_i = \alpha + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i$ 中的 ε_i 总是服从一个期望为0，方差为 σ^2 的典型独立分布。但究竟是何种具体分布是不知道的。现在我们专门研究当 ε_i 服从正态分布时，即 $\varepsilon_i \sim i.i.d\ N(0,\sigma^2)$，多元线性回归模型 $y_i = \alpha + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i$ 的最小二乘估计量 $\hat{\beta}_1$ 的性质。

在问题展开前，先做一下准备工作，复习一下有关正态分布的那些分布。

主要有如下四条结论：

1. 如果 N_i 服从正态分布，那么它的线性组合 $\sum_{i=1}^{n} a_i N_i$ 也服从正态分布。

$$即：\sum_{i=1}^{n} a_i N_i \sim N(\mu,\sigma^2)$$

2. 如果 N_i 服从标准正态分布，那么它的平方和将服从 χ^2 分布。

$$即：\sum_{i=1}^{n} N_i{}^2 \sim \chi_n^2$$

3. 如果 N 服从标准正态分布，Z 服从 χ_n^2 分布，那么 $\frac{N}{\sqrt{Z/n}}$ 服从 t 分布。

$$即：\frac{N}{\sqrt{Z/n}} \sim t_n$$

4. 如果 Z_1 服从 χ_{n1}^2 分布，Z_2 服从 χ_{n2}^2 分布，那么 $\frac{Z_1/n_1}{Z_2/n_2}$ 服从 F_{n_1,n_2} 分布。

$$即：\frac{Z_1/n_1}{Z_2/n_2} \sim F_{n_1,n_2}$$

具体的分析，如果读者感兴趣，可以复习一下前面所讲的内容。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第一节　问题的引入

对于数据生成过程 $y_i = \alpha + \beta x_i + \varepsilon_i, \varepsilon_i \sim i.i.N(0,\sigma^2)$ 来说（注意：从现在起，我们都在正态分布条件下讨论问题，且 y_i 与 ε_i 一样都是正态分布），会有 $y_i \sim i.i.d\ N(\alpha + \beta x_i, \sigma^2)$。前面我们已经证明过对模型 $y_i = \alpha + \beta x_i + \varepsilon_i$ 有 $\hat{\beta} = \sum \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2} \cdot y_i = \sum w_i y_i (w_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2})$。

因此，$\hat{\beta}$ 是关于 y_i 的线性组合，而 $\hat{\beta}$ 服从正态分布，由于正态分布的线性组合仍是正态分布，所以 $\hat{\beta}$ 服从正态分布。我们已知道 $\hat{\beta}$ 的期望无偏，即 $E(\hat{\beta}) = \beta$。$\hat{\beta}$ 的方差为 $\frac{\sigma^2}{\sum (x_i - \bar{x})^2}$。因此 $\hat{\beta}$ 分布的具体形态如下：$\hat{\beta} \sim N(\beta, \frac{\sigma^2}{\sum (x_i - \bar{x})^2})$。

同理，$\hat{\alpha} \sim N[\alpha, (\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2})\sigma^2]$。

也就是说，当 $\varepsilon_i \sim i.i.N(0,\sigma^2)$ 时，$\hat{\alpha}$、$\hat{\beta}$ 分布也为正态分布。问题在于在实际中我们很可能并不知道 σ^2 具体的值，这样也就无法对 $\hat{\alpha}$、$\hat{\beta}$ 进行区间估计。

第二节　问题的预解决

既然不知道 σ^2 的值，能不能找一个估计量代替 σ^2 呢？这是解决问题的起点。

对数据生成过程：$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} \cdots + \beta_k x_{ki} + \varepsilon_i$ 来说，定义估计量 $S^2 = \frac{1}{n-k-1} \sum \hat{\varepsilon}_i^2 = \frac{ESS}{n-k-1}$。

S^2 具有如下性质：

(1) $E(S^2) = \sigma^2$；(2) $(n-k-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-k-1}$ 其中 k 为变量个数。

第一个性质由 S^2 的定义保证，第二个性质的论证思路大致如下：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - (\hat{\alpha} + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki})$$

由于 y_i 本身为正态分布，又因为 $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \cdots \hat{\beta}_k$ 也为正态分布，$y_i - \hat{y}_i$ 这一线性组合也就当之无愧服从正态分布了。也就是说，如果 ε_i 服从正态分布 $N(\mu, \sigma^2)$，那么 $\sum \hat{\varepsilon}_i^2$ 服从 χ^2 分布。

将 ε_i 标准化，引入新变量 N_i : $N_i = \dfrac{\hat{\varepsilon}_i - 0}{\sigma} \sim N(0,1)$。那么，$\sum_{i=1}^{n} N_i^{\ 2} = \dfrac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\sigma^2} \sim \chi^2_{n-k-1}$。

再一次变形：$\dfrac{(n-k-1) \cdot [\dfrac{1}{(n-k-1)} \cdot \sum \hat{\varepsilon}_i^2]}{\sigma^2} \sim \chi^2_{n-k-1}$。

根据定义：$S^2 = \dfrac{\sum \hat{\varepsilon}_i^2}{n-k-1}$，可知 $(n-k-1)\dfrac{S^2}{\sigma^2} \sim \chi^2_{n-k-1}$。

第三节　派生的内容：自由度

从前面的讨论中得知，$(n-k-1)\dfrac{S^2}{\sigma^2}$ 服从 χ^2_{n-k-1} 分布，其自由度为：$n-k-1$，其中 n 表示的是样本个数，k 为方程个数。一般来说，自由度表示方程中能够自由变动的变量个数。

例如：联立方程 $\begin{cases} y_1 + y_2 + y_3 = 7 \\ y_1 = 7 \end{cases}$

在这个方程组中，我们只要在 y_2，y_3 中再任意固定一个，另一个就会求出来。也就是说，y_2，y_3 中只有一个是可以自由变动的变量，这个方程组的自由度为 1。

再看：$\begin{cases} y_1 + y_2 + y_3 + y_4 = 7 \\ y_1 = 7 \end{cases}$

这个方程组中有两个方程，四个变量。很明显，只要我们固定了 y_2，y_3，y_4 中任意两个变量，就可以求出第 3 个。即这里的三个未定变量中只要固定其中两个就行了，因此，该方程组的自由度为 2。

一般来说，我们可以给出自由度公式如下：

$$\text{自由度} = \text{变量个数} - \text{方程个数}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

具体到 $n-k-1$，这个自由度是怎么来的呢？

最小二乘法实际上是求以下函数的极值：

$$\underset{\alpha,\beta_i}{\mathrm{Min}}\sum(y_i-\hat{\alpha}-\hat{\beta}_1x_{1i}-\cdots-\hat{\beta}_kx_{ki})^2$$

求导可知：

$$\begin{cases}\sum(y_i-\alpha-\beta_1x_{1i}-\cdots-\beta_kx_{ki})=0(\text{对}\ \alpha\ \text{求导})\\\sum x_{1i}(y_i-\alpha-\beta_1x_{1i}-\cdots-\beta_kx_{ki})=0(\text{对}\ \beta_1\ \text{求导})\\\vdots\\\sum x_{ki}(y_i-\alpha-\beta_1x_{1i}-\cdots-\beta_kx_{ki})=0(\text{对}\ \beta_k\ \text{求导})\end{cases}$$

可以变形为

$$\begin{cases}\sum\varepsilon_i=0\\\sum x_{1i}\varepsilon_i=0\\\sum x_{2i}\varepsilon_i=0\\\vdots\\\sum x_{ki}\varepsilon_i=0\end{cases}$$

展开后有：

$$k+1\ \text{个方程}\overbrace{\begin{cases}\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_n=0\\x_{11}\varepsilon_1+x_{12}\varepsilon_2+\cdots+x_{1k}\varepsilon_k+\cdots+x_{1n}\varepsilon_n=0\\\vdots\\x_{ki}\varepsilon_1+x_{k2}\varepsilon_2+\cdots+x_{kk}\varepsilon_k+\cdots+x_{kn}\varepsilon_n=0\end{cases}}^{n\text{个变量}}$$

这里一共 n 个变量 ε_i，$k+1$ 个方程。因此该方程组的自由度为 $n-(k+1)=n-k-1$，这里你也许会问，怎么有 n 个变量 ε_i 呢？回过来看看下面这张表，你就会清楚了：

$$\begin{matrix}y_1 & x_{11} & x_{12} & \cdots x_{1k} & \varepsilon_1\\y_2 & x_{21} & x_{22} & \cdots x_{2k} & \varepsilon_2\\\vdots & \vdots & \vdots & \vdots & \vdots\\y_n & x_{n1} & x_{n2} & \cdots x_{nk} & \varepsilon_n\end{matrix}$$

自由度 $n-k-1$ 告诉我们，每 n 个 ε_i 中，只有 $n-k-1$ 个为自由变量，余下的 $k+1$ 个可通过上面的方程组求出。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第四节 回归系数的假设检验

先看 $\hat{\beta}$ 的分布：如果 σ^2 已知，那么 $\hat{\beta} \sim N[\beta, \frac{\sigma^2}{\sum (x_i - \bar{x})^2}]$。

如果 σ^2 不知道呢？（实际上 σ^2 往往是未知的）

由于 $N = \frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum (x_i - \bar{x})^2}}}$ 服从 N（0，1）；$Z = \frac{(n-k-1)S^2}{\sigma^2}$ 服从 χ^2_{n-k-1}，那么，$t =$

$$\frac{N}{\sqrt{\frac{Z}{n-k-1}}} = \frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum (x_i - \bar{x})^2}}} \Big/ \sqrt{\frac{S^2}{\sigma^2}} = \frac{\hat{\beta} - \beta}{\sqrt{\frac{S^2}{\sum (x_i - \bar{x})^2}}} \sim t_{n-k-1}。$$

例如，对 $y_1 = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$，$\varepsilon_i$ 服从正态分布，$E(\varepsilon_i) = 0$，但不知方差，请检验 $H_0 : \beta_1 = 0$。

令 $t_0 = \frac{\hat{\beta}_1 - 0}{\sqrt{\frac{S^2}{(1 - R_1^2) \sum (x_i - \bar{x})^2}}} \sim t_{n-k-1}$，其中 $S^2 = \frac{ESS}{n-k-1}$。

如果所计算出的 t_0 在临界值外，即 t_0 很大，就拒绝原假设 $H_0 : \beta_1 = 0$，这说明 x_{1i} 对方程有显著影响，我们称估计量 $\hat{\beta}_1$ 显著。

如果 t_0 在临界值内，则接受原假设 $H_0 : \beta_1 = 0$，这说明 x_{1i} 对方程无显著影响，我们通常称估计量 $\hat{\beta}_1$ 不显著。

在计算机输出的结果中，所谓估计量的 t 检验值，指在原假设 $H_0 : \beta_1 = 0$ 时的 t 检验值，因此，一般来说，检验值越大越好，该检验值越大，拒绝假设 $H_0 : \beta_1 = 0$ 越肯定。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第五节 预 测

前面早已说过，对于不同的样本，会回归出不同的拟合直线。当然，拟合直线本身变化也有一定范围。这样，就引导我们讨论有关预测的问题。利用 $y_i = \alpha + \beta x_i$ 这一拟合直线进行预测时有两种设想：第一种，不考虑扰动项 ε_i，仅对 $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$ 的均值进行预测研究，而 $E(\alpha + \beta x_i)$ 恰恰又是 $y_i = \hat{\alpha} + \hat{\beta}x_i + \varepsilon_i$ 的均值 $E(\hat{y}_i)$。因此，我们把它叫做一般水平的均值预测。第二种，考虑了扰动项 ε_i 对 $y_i = \hat{\alpha} + \hat{\beta}x_i + \varepsilon_i$ 进行预测研究。我们把这种预测称为个体水平的均值预测。在第一种情况下，由于实际的产生过程中有 ε_i 的影响，必然对每一个 x_i 有多种 y_i 值对应，那么对 y_i 的每一个估计值来说，由于没有考虑 ε_i，其方差会相应地小些。而第二种情况实际等于给定一个 y_i，要求出 y_i 的变化范围，故而要考虑 ε_i，所以这类方差通常要大一些。

一、一般水平的均值预测（$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$）

早已说过，对于 $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$，很明显，它服从正态分布，其期望为 $E(\hat{y}_i)$，方差为

$$
\begin{aligned}
Var(\hat{y}_i) &= Var(\hat{\alpha} + \hat{\beta}x_i) \\
&= Var(\hat{\alpha}) + Var(\hat{\beta}x_i) + 2\mathrm{cov}(\hat{\alpha}, \hat{\beta}x_i) \\
&= \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}\right)\sigma^2 + x_i^2 \frac{\sigma^2}{\sum (x_i - \bar{x})^2} + 2x_i\mathrm{cov}(\hat{\alpha}, \hat{\beta}) \text{ ④} \\
&= \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}\right)\sigma^2 + \frac{x_i^2\sigma^2}{\sum (x_i - \bar{x})^2} + 2x_i \frac{-\bar{x}}{\sum (x_i - \bar{x})^2}\sigma^2 \text{ ⑤} \\
&= \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]\sigma^2
\end{aligned}
$$

可见 $\hat{y}_i \sim N\left[E(\hat{y}_i), \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)\sigma^2\right]$。

---

④ 注意，这里对于一个具体的 x_i 都有具体值对应。x_i 是个常量，所以能够提到 cov（$\hat{\alpha}$，$\hat{\beta}$）括号外。

⑤ cov（α，β）的展开式结论可以在第三章有关 $y_i = \alpha + \beta x_i + \varepsilon_i$ 的研究中找到。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



将其标准化：$\dfrac{\hat{y}_i - E(\hat{y}_i)}{\sqrt{\sigma^2[\dfrac{1}{n} + \dfrac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}]}} \sim N(0,1)$ 。

自然，在 σ^2 未知时，用 σ^2 的无偏估计量 S^2 代替，那么，上式变为：

$$\frac{\hat{y}_i - E(\hat{y}_i)}{\sqrt{S^2[\dfrac{1}{n} + \dfrac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}]}} \sim t_{n-k-1}$$

所有这一切，都是告诉我们，$\hat{y}_i$ 是可以算出的，同时，只要给一个置信度，就可以找出 $E(\hat{y}_i)$ 的范围。因此，给定一个 x_i，一个置信度 α，就可以算出 $E(\hat{y}_i)$ 的方差 $t_{\frac{\alpha}{2}}S^2[\dfrac{1}{n} + \dfrac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}]$，由于存在开方，必然有对称，同时，$S^2[\dfrac{1}{n} + \dfrac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}]$ 也告诉我们这是一条双曲线。画出它的图形如图 6.1 所示。

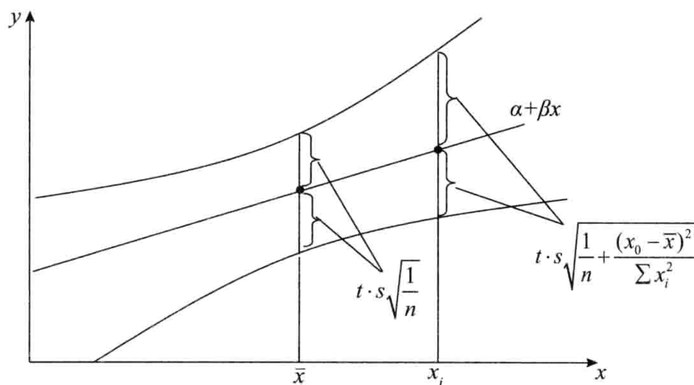


图 6.1

这是一条以 $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$ 为轴的双曲线。两线所夹范围为 $\hat{y}_i$ 的方差，当 $x_i = \bar{x}$ 时，$\hat{y}_i$ 的方差为 $\dfrac{1}{n}S^2$，达到最小。随着 x_i 偏离 $\bar{x}$，方差会越来越大。也就是说，方差在 $\pm S^2[\dfrac{1}{n} + \dfrac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}]$ 范围里。这是因为 $\sum (x_i - \bar{x})^2$ 是可变的，但随着 x_i 偏离 $\bar{x}$ 的程度加大。$\sum (x_i - \bar{x})^2$ 不断增大，即 $\hat{y}_i$ 的方差完全由 $\dfrac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}$ 和样本个数 n 决定。因此，上面

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

的分析告诉我们，对给定的一个 x_i 越靠近现有的 $\bar{x}$，对一般水平的均值预测就会越准确。

二、个体水平的均值预测（$\tilde{y}_i = \hat{\alpha} + \hat{\beta}x_i + \varepsilon_i$）

在使用 $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$ 预测 $E(\hat{y}_i)$ 的变化时，没考虑到 ε_i 的作用。现在，把 ε_i 考虑进去，可使用估计量 $\tilde{y}_i = \hat{\alpha} + \hat{\beta}x_i + \varepsilon_i$，很显然 $\tilde{y}_i$ 也同样服从正态分布，其期望是 $E(\tilde{y}_i)$，但其方差为：$Var(\tilde{y}_i) = Var(\varepsilon_i) + Var(\hat{\alpha} + \hat{\beta}x_i) + 2\text{cov}(\hat{\alpha} + \hat{\beta}x_i, \varepsilon_i)$（这里又要用到我们在第三章中学过的最小二乘估计量性质 $\text{cov}(\hat{\alpha} + \hat{\beta}x_i, \varepsilon_i) = 0$）。

$$
\begin{aligned}
Var(\tilde{y}_i) &= Var(\varepsilon_i) + Var(\hat{\alpha} + \hat{\beta}x_i) \\
&= \sigma^2 + \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]\sigma^2 \\
&= \left[1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]\sigma^2
\end{aligned}
$$

因此，$\tilde{y}_i \sim N\left[E(\tilde{y}_i), \left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)\sigma^2\right]$。

进行标准化：

$$
\frac{\tilde{y}_i - E(\tilde{y}_i)}{\sqrt{\sigma^2\left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)}} \sim N(0,1)
$$

如果 σ^2 未知，那么，将 σ^2 的无偏估计量 S^2 代入上式，则

$$
\frac{\tilde{y}_i - E(\tilde{y}_i)}{\sqrt{S^2\left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)}} \sim t_{n-k-1}
$$

记住个体水平下的均值预测，就是说，对每一个 x_i，在考虑 ε_i 的情况下，y_i 的变动范围是什么？由于个体水平下的均值预测考虑了 ε_i，因此 y_i 的方差：$S^2\left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)$。当然，它比不考虑 ε_i 时的 $\hat{y}_i$ 的方差要大一些。道理很简单，后者没有考虑扰动项，而前者考虑了扰动项，方差自然就大上去了。

最后应该指出，考虑 ε_i 的影响是关键的一环，不考虑它，就是一般水平的均值预测：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$\hat{y}_i \sim N[E(\hat{y}_i),(\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2})\sigma^2]$；考虑它，方差自然要大一些了，$\tilde{y}_i \sim N[E(\tilde{y}_i),(1+\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2})\sigma^2]$。

第六节 复习与提高

本章仍然主要讨论 $y_i=\alpha+\beta x_{1i}+\varepsilon_i$，这种一元线性回归模型的问题。但 ε_i 的分布具体到了正态分布。所有的讨论自然都是在正态分布下展开的。

$\hat{\alpha}$、$\hat{\beta}$ 的分布：

$$\hat{\alpha} \sim N[\alpha,(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2})\sigma^2]$$

$$\hat{\beta} \sim N[\beta,\frac{1}{\sum(x_i-\bar{x})^2}\sigma^2]$$

定义 $S^2=\frac{1}{n-k-1}\sum\hat{\varepsilon}_i^2$，$S^2$ 又引出了自由度问题，记住了“变量个数 − 方程个数 = 自由度”，用这一公式可使你省不少心。当这一切都完成之后，就可以用区间估计方法找到 β 的区间了。

在一元线性回归模型 $y_i=\alpha+\beta x_{1i}+\varepsilon_i$ 中，共有两种估计量：α、β 的点估计量；区间估计量。

最后，我们讨论有点预测的问题。当然考虑 ε_i 的影响是关键的一环，不考虑它，就是一般水平的均值预测：$\hat{y}_i \sim N[E(\hat{y}_i),(\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2})\sigma^2]$；考虑它，方差自然要大一些了，$\tilde{y}_i \sim N[E(\tilde{y}_i),(1+\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2})\sigma^2]$。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第七章 假设检验

对估计量进行检验，有时是有利于计算简化的好法子。举例来说，对于 $y_i = \alpha + \beta x_i + \varepsilon_i$，如果我们检验出假设 $H_0: \beta = 0$ 成立，这说明 x_i 的变化对 y_i 本身丝毫不起作用，这必将使我们对模型的形式进行转变。而且，我们的估计量总是有误差存在的，因此，有必要对估计量进行检验，本章讨论的是假设检验的一般方法和技巧。

第一节 t 检验：对单个参数检验的方法

对于数据生成过程：$y_i = \alpha + \beta x_i + \varepsilon_i$ 来说，用最小二乘法求得的 $\hat{\beta}$ 服从 $N[\beta, \frac{\sigma^2}{\sum (x_i - \bar{x})^2}]$ 正态分布。当不知道 σ^2 具体值时，用 S^2 代替 σ^2，但要用 t 分布来代替正态分布了，因此，$\frac{\hat{\beta} - \beta}{\sqrt{S^2 \frac{1}{\sum (x_i - \bar{x})^2}}} \sim t_{n-k-1}$。

一、检验假设 $H_0: \beta = 0$

对假设 $H_0: \beta = 0$ 进行 t 检验的步骤如下：

1. 我们利用已知数据算出 $\hat{\beta}$；

2. 再由 $S^2 = \frac{ESS}{n-k-1}$，算出 $\sqrt{\frac{S^2}{\sum (x_i - \bar{x})^2}}$；

3. 计算统计量 $t_0 = \frac{\hat{\beta} - \beta}{\sqrt{\frac{S^2}{\sum (x_i - \bar{x})^2}}}$，代入假设 $H_0: \beta = 0$，则 $t_0 = \frac{\hat{\beta}}{\sqrt{\frac{S^2}{\sum (x_i - \bar{x})^2}}}$；

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



4. 根据所给出的置信度 α，查表找出与其相应的 t 检验的临界值 $t_{\frac{\alpha}{2}}$；

5. 作出统计推断：如果 t_0 大于临界值 $t_{\frac{\alpha}{2}}$，就拒绝 $\beta=0$；否则接受 $\beta=0$，说明 x_i 对 y_i 无任何影响。

从上面的分析可以看出，我们一般希望 t_0 的值越大越好。t_0 的值越大，x_i 对 y_i 的作用越显著。因此，t_0 叫做**显著水平**（Level of Significance），它越大越理想。

一般而言，计算机输出的结果中的 t 检验值，隐含着 $H_0: \beta=0$ 原假设。

实际中我们有时只需看一下所计算出的统计量 t_0，便知取舍。如果 t_0 大到 5 以上，我们便根本不用考虑查表，就说明很好，如果 t_0 等于 2 点多，就得查一查表了。

二、t 检验的技巧转换

有时往往不需要直接的 t 检验值，只想弄清估计量之间的关系如何。比如 $H_0: \alpha=\beta$，$H_0: \alpha+5=\beta+4$ 等，这种情况下，怎么使用 t 检验对原假设进行检验呢？这不像我们刚接触的 t 检验假设时是单个参数。由此我们就不妨先造出一个参数来，如引入一个新变量 $\hat{\theta}$，让 $\hat{\theta}$ 等于原假设 H_0 左右两边之差。接下来就是检验 $\hat{\theta}$ 是不是等于 0 的问题了。

举例来说，对一元线性回归模型：$y_i=\alpha+\beta x_i+\varepsilon_i$，要对假设 $H_0: \alpha=\beta$ 进行检验。

1. 首先要令：$\hat{\theta}=\hat{\alpha}-\hat{\beta}$

2. 由于 $\hat{\alpha}$、$\hat{\beta}$ 都是正态分布，$\hat{\theta}$ 当然服从正态分布，则

$$E(\hat{\theta})=E(\hat{\alpha}-\hat{\beta})=E(\hat{\alpha})-E(\hat{\beta})=\alpha-\beta$$

$$\begin{aligned}Var(\hat{\theta}) &= Var(\hat{\alpha}-\hat{\beta})\\ &= Var(\hat{\alpha})+Var(\hat{\beta})-2\mathrm{cov}(\hat{\alpha},\hat{\beta})\\ &= \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right)+\frac{\sigma^2}{\sum(x_i-\bar{x})^2}+\frac{2\bar{x}\sigma^2}{\sum(x_i-\bar{x})^2}\\ &= \sigma^2\left[\frac{1}{n}+\frac{(1+\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]\end{aligned}$$

所以，$\hat{\theta} \sim N\left(\alpha-\beta,\sigma^2\left[\frac{1}{n}+\frac{(1+\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]\right)$。

标准化后：$$\frac{\hat{\theta}-\theta}{\sqrt{\sigma^2\left[\frac{1}{n}+\frac{(1+\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]}} \sim N(0,1)$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

3. 计算统计量 t_0

现在原假设 $H_0: \alpha = \beta$ 就变成了 $H_0: \theta = \alpha - \beta = 0$，在未知 σ^2 情况下，则

$$t_0 = \frac{\hat{\theta}}{\sqrt{S^2\left[\frac{1}{n} + \frac{(1+\bar{x})^2}{\sum (x_i - \bar{x})^2}\right]}} \sim t_{n-k-1}$$

4. 做出统计推断

当 $|t_0| \geq t_{\frac{\alpha}{2}}$ 很大时，拒绝假设 $H_0: \theta = 0$，说明 $\alpha \neq \beta$；

当 $|t_0| \leq t_{\frac{\alpha}{2}}$ 很小时，接受假设 $H_0: \theta = 0$，说明 $\alpha = \beta$。

第二节　F 检验的引入

在假设检验中诚然能运用引入 θ 变量的技巧，但这种方法毕竟有一定局限性，有时会给计算带来很大又不必要的麻烦，况且，t 检验仅仅限于对单个假设进行检验。将要引入的 F 检验不受这种限制，它可在一般情况下对多个假设进行检验。

对于数据生成过程 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$，原假设 H_0 可能包括多个假设，比如：$H_0: \begin{cases} \beta_1 = \beta_2 \\ \beta_3 = 5 \end{cases}$

现在我们来看 F 检验的思路：

第一步，如果假设是对的，代入到原方程中去：

$$y_i = \alpha + \beta_1(x_{1i} + x_{2i}) + 5x_{3i} + \cdots + \beta_k x_{ki} + \varepsilon_i$$

整理后，可得

$$(y_i - 5x_{3i}) = \alpha + \beta_1(x_{1i} + x_{2i}) + \cdots + \beta_k x_{ki} + \varepsilon_i$$

令 $y_i^* = y_i - 5x_{3i}$，$\alpha = \alpha^*$，$x_{1i}^* = x_{1i} + x_{2i}$，则：$y_i^* = \alpha^* + \beta_1 x_{1i}^* + \cdots + \beta_k x_{ki} + \varepsilon_i$。

第二步，在新的模型中，可以得到新的回归残差平方和 ESS^*；而对于原方程 y_i 也可得到相应的残差平方和 ESS。

如果假设是对的，ESS^* 应该接近或等于 ESS，即（$ESS^* - ESS$）应该很小很小。为了方便起见，改用一个相对量后，即

$$F_0 = \frac{ESS^* - ESS}{ESS}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



前面已经说过 ESS 服从 χ^2 分布，F_0 将服从 F 分布。因此，问题的焦点就集中在 F 分布上了。

第三节 一般假设检验：F 检验

前面，我们已经知道指标 $F_0 = \dfrac{ESS^* - ESS}{ESS}$ 是 F 检验的关键，但是关键的关键在于 ESS 的特征，只要弄清楚 ESS 的分布后，一切问题都会迎刃而解。

下面，我们就讨论一下有关 ESS 的分布问题。

$$ESS = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_k x_{ki})^2$$

由于 y_i、$\hat{y}_i$ 都是正态分布，因此，$\hat{\varepsilon}_i = y_i - \hat{y}_i$ 服从 $N(0,\sigma^2)$ 的正态分布，自然 $(y_i - \hat{y}_i) \sim N(0,\sigma^2)$。

标准化之后，可得：$\dfrac{y_i - \hat{y}_i}{\sigma} \sim N(0,1)$；再求平方和：$\dfrac{\sum \hat{\varepsilon}_i^2}{\sigma^2} = \dfrac{\sum (y_i - \hat{y}_i)^2}{\sigma^2}$。

从 χ^2 分布与正态分布的关系，可知：$\dfrac{\sum (y_i - \hat{y}_i)^2}{\sigma^2} \sim \chi^2_{n-k-1}$，即 $\dfrac{ESS}{\sigma^2} \sim \chi^2_{n-k-1}$。

这个关键的结打开之后，会给下面的讨论带来很多方便。

下面讨论 F 检验的具体步骤：

1. 计算未受约束的回归残差平方和

对于数据生成过程 $y_i = \alpha + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i$，我们能够得到一个未受约束的回归残差平方和，记成 ESS(Unrestricted) 或 $ESS(u)$，它的自由度为 $n-k-1$。

2. 计算约束回归残差平方和

把要检验的等式代入未受约束回归模型中去，形成新方程，并以此进行回归。

我们仍使用在介绍 F 检验思路时使用的例子，将原假设 H_0 代入未受约束回归模型，可得

$$y_i^* = \alpha^* + \beta_1 x_{1i}^* + \beta_4 x_{4i} + \cdots + \beta_k x_{ki}$$

利用新方程可得到约束条件下的回归残差平方和，记成 ESS(Restricted) 或 ESS(R)，自由度将变成 $n-(k-2)-1$，自由度增大。这是因为 x_{2i} 合并到了 x_{1i} 中，x_{3i} 合并到了 y_1^* 中。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

这里把它作为规律介绍给你：**假设检验 H_0 中所包括的约束条件的个数，就是减少的变量的个数**。或者，你干脆就数一数假设 H_0 中等号的个数，等号的个数就是减少变量的个数。

新方程的新自由度 $= n-(k-r)-1$，其中：r 为约束条件的个数，或假设 H_0 中等号的个数。

3. 约束的成本

新得到的受约束的回归残差平方和 $ESS(R)$，由于减少了变量个数，使得它必然大于未受约束的残差平方和 $ESS(u)$ 。我们可以定义：

$$\text{约束成本} = ESS(R) - ESS(u)$$

$$\begin{aligned}\text{约束成本的自由度} &= ESS(R)\text{ 的自由度} - ESS(u)\text{ 的自由度} \\ &= [n-(k-r)-1]-(n-k-1) \\ &= r\end{aligned}$$

4. 相对约束成本

仅仅依靠约束成本的绝对量一般很难判断出假设 H_0 的优劣程度，采用相对量可以帮助我们解决这一问题。

$$\text{相对成本} = \frac{ESS(R) - ESS(u)}{ESS(u)}$$

5. F-检验

首先，计算估计量 $F_0 = \dfrac{[ESS(R) - ESS(u)]/r}{ESS(u)/n-k-1} \sim F_{r,n-k-1}$ ；再根据所计算出的 F_0 和置信度 α 对原假设 H_0 做出统计推断。

当然，我们还可以采用另外一种方法来计算 F_0，具体方法如下：

由于 $ESS(u) = TSS_u(1-R_u^2)$，$ESS(R) = TSS_R(1-R_R^2)$ ；如果 $TSS_R = TSS_u$，

$$\text{则：} F_0 = \frac{TSS[(1-R_R^2)-(1-R_u^2)]/r}{[TSS(1-R_u^2)]/n-k-1}$$

$$= \frac{(R_u^2 - R_R^2)/r}{(1-R_u^2)/(n-k-1)}$$

但是使用上式时，有个条件必须小心，即 $TSS_R = TSS_u$，这意味着 $y_i^* = y_i$ 是先决条件。在给出 R_R^2 或 R_u^2 较易求出时，使用这个公式效果不错。比如，我们进行的实际回归操作中，对下面方程中的假设 $H_0: \beta_2 = 0$ 进行检验，用这个变形公式就很容易：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad R_u^2 = 0.997$$

$$y_i = \alpha + \beta_1 x_{1i} + \varepsilon_i \qquad R_R^2 = 0.375$$

$$F_0 = \frac{(0.997 - 0.375)/1}{(1 - 0.997)/(10 - 2 - 1)} > F_{1,7}$$

因此，拒绝假设 $H_0: \beta_2 = 0$。

当然，这检验也可以用 $F_0 = \frac{[ESS(R) - ESS(u)]/r}{ESS(u)/n-k-1}$ 来进行。

第四节　F 检验的附加例子

一、检验回归系数的等价性

例如对假设 $H_0: \beta_i = \beta_j$ 进行检验，其具体步骤如下：

1. 进行未受约束回归：$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_j x_{ji} + \cdots + \beta_k x_{ki} + \varepsilon_i$；
2. 进行约束回归：$y_i = \alpha + \beta_1 x_{1i} + \cdots + \beta_i (x_{ii} + x_{ji}) + \cdots + \beta_k x_{ki} + \varepsilon_i$；
3. 进行 F 检验 $F_0 = \frac{[ESS(R) - ESS(u)]/1}{ESS(u)/n-k-1} \sim F_{1,n-k-1}$。

当然，如果 $H_0: \beta_k = \beta_m, \beta_i = \beta_j, \beta_L = \beta_n, \cdots$，假设的等号增加了。检验值 F_0 分子的数字也随之增加，$ESS(R)$ 也发生变化，具体方法与检验 $H_0: \beta_i = \beta_j$ 的方法相同。

二、检验回归系数的线性组合

例如对假设 $H_0: \beta_i = \frac{1}{3}\beta_k + \frac{4}{3}\beta_{i+1}$ 进行检验，其具体步骤如下：

1. 进行未受约束回归：$y_i = \alpha + \beta_1 x_{1i} + \cdots + \beta_i x_{ii} + \beta_{i+1} x_{i+1,i} + \cdots + \beta_k x_{ki} + \varepsilon_i$；
2. 进行受约束回归：$y_i = \alpha + \beta_1 x_{1i} + \cdots + (\frac{1}{3}\beta_k + \frac{4}{3}\beta_{i+1}) x_{ii} + \beta_{i+1} x_{i+1,i} + \cdots + \beta_k x_{ki} + \varepsilon_i$；

那么，$y_i = \alpha + \beta_1 x_{1i} + \cdots + \beta_{i+1} (\frac{4}{3} x_{ii} + x_{i+1,i}) + \cdots + \beta_k (\frac{1}{3} x_{ii} + x_{ki}) + \varepsilon_i$。

令 $x_{i+1,i}^* = \frac{4}{3} x_{ii} + x_{i+1,i}$，$x_{ki}^* = \frac{1}{3} x_{ii} + x_{ki}$，则约束回归模型为：$y_i = \alpha + \beta_1 x_{1i} + \cdots + \beta_{i+1} x_{i+1,i}^* + \cdots + \beta_k x_{ki}^* + \varepsilon_i$。

3. 计算统计量 F_0，并以此进行统计推断。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$F_0 = \frac{[ESS(R) - ESS(u)]/1}{ESS(u)/n-k-1} \sim F_{1,n-k-1}$$

当然，也可以用公式 $F_0 = \frac{(R_u^2 - R_R^2)/r}{(1-R_u^2)/(n-k-1)}$ 来计算 F_0，在这个例子中，TSS_u 与 TSS_R 是相等的，符合条件。

三、常数检验

例如对假设 $H_0: \beta_2 = 5$ 进行检验，其具体步骤如下：

1. 进行未约束回归：$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$

$$y_i = \alpha + \beta_1 x_{1i} + 5x_{2i} + \varepsilon_i$$

2. 进行约束回归：$\Rightarrow y_i - 5x_{2i} = \alpha + \beta_1 x_{1i} + \varepsilon_i$

$$\Rightarrow y_i^* = \alpha + \beta_1 x_{1i} + \varepsilon_i (y_i^* = y_i - 5x_{2i})$$

3. 计算统计量 F_0，并以此进行统计推断

这里不能使用 $F_0 = \frac{(R_u^2 - R_R^2)/r}{(1-R_u^2)/(n-k-1)}$ 来计算 F_0 了，因为 TSS 变了，只好用 ESS 来计算了。

四、检验连等式

例如对假设：$H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0$ 进行检验。

这是一个包括 k 个等号或 k 个约束的假设，对于 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$，而言，如果该假设 H_0 成立，则证明了所有的 x_i 将变得毫无意义，y_i 将丝毫不受所有 x_i 的影响。

该假设的约束模型为：$y_i = \alpha + \varepsilon_i$

那么，$ESS(R) = \sum (y_i - \hat{y}_i)$

$$= \sum (y_i - \hat{\alpha})^2 \text{（因为 } y_i = \alpha + \varepsilon_i \text{ 的最小二乘估计量为：} \hat{\alpha} = \bar{y}\text{）}$$

$$= \sum (y_i - \bar{y})^2$$

$$= TSS(u)$$

$$F_0 = \frac{[ESS(R) - ESS(u)]/k}{ESS(u)/n-k-1} = \frac{[TSS(u) - ESS(u)]/k}{ESS(u)/n-k-1}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



一般来说，在计算机输出结果中所给出的 F 检验值是在假设 $H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0$ 下的 F 检验值。这个 F_0 检验值力图描述所有 x_i 的整体对 y_i 的影响。

值得注意的是 $ESS(u)$ 的自由度为 $n-k-1$，$TSS(u)$ 的自由度却无论什么时候都是 $n-1$。这是因为 $TSS = \sum (y_i - \bar{y})^2$，而 $\bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n}$ 有 n 个变量，那么，TSS 的自由度为：变量个数 - 方程个数 $= n-1$，即 TSS 的自由度为 $n-1$。

第五节　调整后的相关系数 $\bar{R}^2$

相关系数 $R^2 = 1 - \frac{ESS}{TSS}$，我们早已烂熟于胸。但这个 R^2 没有考虑自由度，故使得变量越多，R^2 会越大。如图 7.1 所示，即随着变量的增加，R^2 会无限接近于 1。

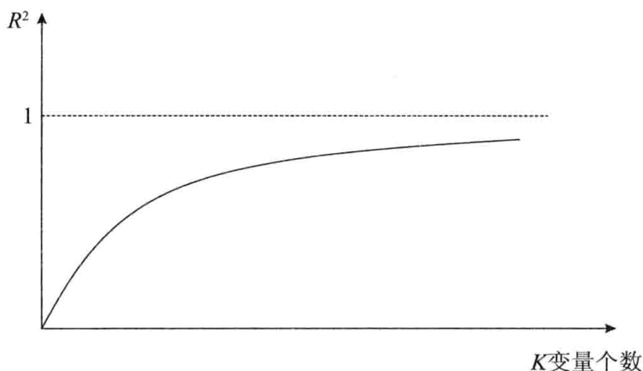


图 7.1

如果把自由度考虑到 R^2 中去，则 R^2 将变成 $1 - \frac{ESS(u)/n-k-1}{TSS/(n-1)}$，这是瑟耳（Theil）最早提出来的，为了与通常意义上的 R^2 区别起见，称为瑟耳的 R^2，或调整后的 R^2，记作 $\bar{R}^2$。

$$\bar{R}^2 = 1 - \frac{ESS(u)/n-k-1}{TSS/n-1} = 1 - \left(\frac{n-1}{n-k-1}\right) \cdot (1 - R^2)$$

由于考虑了自由度，$\bar{R}^2$ 并不简单地随着变量个数的增加而增大，当变量个数增加到一定程度时，它将开始下倾，如图 7.2 所示。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

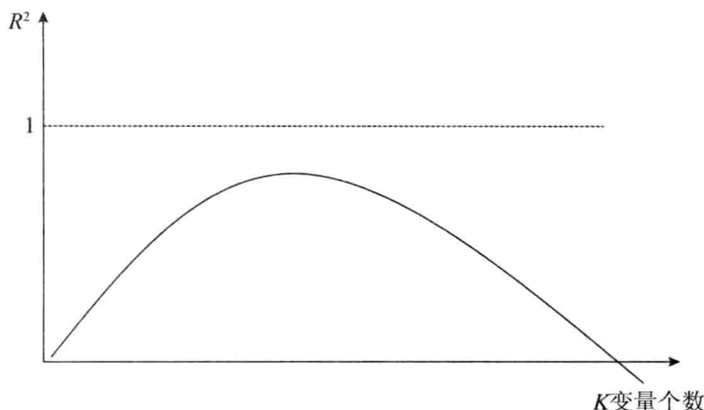


图 7.2

$\bar{R}^2$ 是如何运动的呢？由于 $\bar{R}^2 = 1 - \frac{n-1}{n-k-1} \cdot (1 - R^2)$，随着变量个数的增加，有两种作用：一是 R^2 增长，R^2 增加会引起 $\bar{R}^2$ 的增大，二是 $\frac{n-1}{n-k-1}$ 会增大，引起 $\bar{R}^2$ 的减少。开始时起主要作用的可能是 $(1 - R^2)$，但当变量个数达到一定程度后，即变量增加到一定程度以后，$(1 - R^2)$ 的作用会越来越小，这是因为变量增加到一定程度后，再增加变量不再会对 R^2 的变化有明显影响，故而 $(1 - R^2)$ 的作用也不明显了；而 $\frac{n-1}{n-k-1}$ 的作用，则开始占主导地位，这种作用操纵整个 $\bar{R}^2$ 的变化，使 $\bar{R}^2$ 逐渐下降。

第六节　因果关系检验

判断一个变量的变化是否是引起另一个变量变化的原因，这是计量经济学中的一个常见的问题。例如，是货币供给的变化引起 GNP 的变化，还是 GNP 和货币供给的变化都是内生决定的。Granger 和 Sims 提出的因果关系检验是解决这类问题的一种方法。读者可参见 C. W. J. Granger 的“用计量模型和方法考察因果联系”⑥ 和 C. A. Sims 的“货币、

⑥ C. W. J. Granger, "Investigating Causal Relations by Econometric Model and Cross - Spectral Methods", Econometrica, V01 37, pp. 424 ~ 438, 1969.

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



收入和因果关系”[7]。有关理性预期的例子，可参见 T. J. Sargent 的“美国古典宏观计量经济模型”[8] 和“理性预期下的动态劳动需求估计”[9]。

其基本思想很简单，如果 X 影响 Y，那么 X 的变化必然应先于 Y 的变化，尤其是要断定“X 影响 Y”时，必须满足两个条件：第一，能够根据 X 预测 Y。也就是说，根据 Y 的过去值对 Y 进行回归时，如果加上 X 的过去值这个因变量，能显著增强回归的解释能力。第二，不能根据 Y 预测 X，因为如果能够根据 X 预测 Y，又能根据 Y 预测 X，很可能 X 和 Y 都是由第三个或更多的其它变量决定的。

为了判断两个条件是否都满足，我们需要对零假设“一个变量不能帮助预测另一个变量”进行检验。例如，要检验“X 不影响 Y”，我们先根据 X 和 Y 的滞后值对 Y 回归（未约束的回归），然后仅用 Y 的滞后值对 Y 进行回归（受约束回归）。于是，我们将采用简单的 F 检验来判断 X 是否显著增强了第一个回归的解释能力。从前文我们可知 F 统计量为：

$$F_0 = \frac{(ESS_R - ESS_u)/r}{ESS_u/n-k-1}$$

其中，ESS_R 和 ESS_u 分别是有约束回归和无约束回归的残差平方和；n 是观察样本容量；k 是无约束回归中变量的个数；r 是约束中的参数个数。这个统计量的分布为 $F(r,n-k-1)$。

如果 X 的滞后值明显有助于增强第一个回归的解释力，我们可以拒绝零假设“X 不影响 Y”，并且可以推得“X 影响 Y”这一结论。同理，对“Y 不影响 X”的假设也可如此检验。

我们采用以下步骤检验 X 是否影响 Y：

首先，进行两个回归检验零假设“X 不影响 Y”。

无约束回归：$Y = \sum_{i=1}^{m} \alpha_i Y_{t-i} + \sum_{i=1}^{m} \beta_i X_{t-i} + \varepsilon_t$，约束回归：$Y = \sum_{i=1}^{m} \alpha_i Y_{t-i} + \varepsilon_t$。

用每个回归的残差平方和计算出 F 统计量并检验系数组 $\beta_1, \cdots, \beta_m$ 是否显著不为零。如果其明显不近似为 0，则可以拒绝“X 不影响 Y”这一假设。

---

⑦ C. A. Sims,"Money, Income, and Causality", America Economic Review, Vol 62, pp. 540 ~ 552, 1972.

⑧ T. J. Sargent,"A classical Macroeconometric Model for the United States", Journal of Political Economy, Vol 84, pp. 207 ~ 238, 1976.

⑨ T. J. Sargent,"Estimation of Dynamic Labor Demand Schedules under Rational Expectations", Journal of Political Economy. V01 86, pp. 1009 ~ 1044, 1978.

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



其次，进行与前面相似的回归检验假设“Y 不影响 X”，只需交换 X 和 Y 的位置并检验 Y 滞后值是否显著不为 0，要得出 X 影响 Y 这一结论，我们必须拒绝假设“X 不影响 Y”，同时接受假设“Y 不影响 X”。

注意回归中的时间间隔 m 的选择是主观的，可归结为一个判断上的问题。通常最好用几个不同的 m 值进行检验，保证不因 m 的选择而不同，同时应注意，因果关系检验的一个缺陷在于实际上可能是由第三个变量 Z 影响 Y，但当期内 Z 与 X 相关。处理这种可能的方法之一就是把 Z 也放在右方进行回归。对 Granger – Sims 因果关系检验的一些批评性验证可参见 R. L. Jacobi，E. E. Leamer 和 M. P. Ward 的“因果关系检验的困难”⑩。和 E. L. Feige 及 D. K. Pearce 的“货币和收入的因果联系：时间序列分析的一些漏洞”⑪。关于因果规律一些更普遍的、近期的讨论，可参见 C. W. J. Granger 的“因果关系概念的一些新发展”⑫ 和 J. W. Pratt 及 R. Schlaifer 的“规律的说明和观测”⑬。

下面通过两个例子来说明因果关系检验的应用：

首先，考察国际石油价格对美国经济的影响。在 20 世纪 70 年代和 80 年代，世界石油价格急剧变化，石油在工业经济中扮演极其重要的角色，这种“石油冲击”有其深远的宏观经济影响。例如，1975 年和 1980 年美国经济衰退的主要原因之一，显然就是 1974 年和 1979—1980 年发生的石油价格暴涨。石油价格的上涨从多方面引起了衰退。首先，石油价格上涨导致了石油进口国的真实国民收入减少。其次，它引起了“调整效应”——通货膨胀的刚性导致的真实收入和产出的进一步下降阻碍了工资和非能源价格迅速达到均衡。关于这些影响的详细讨论，参见 R. S. Pindyck 和 J. J. Rotenberg 的“能源冲击和宏观经济”⑭。

石油价格在 1974 年暴涨前也是起伏波动的。在研究其对宏观经济的影响时，James Hamilton 证明数据资料支持“石油价格导致真实 GNP 及其它主要经济变量变化”这一假

---

⑩ R. L. Jacobi，E. E. Leamer，and M. P. ward，"The Difficuhies with Testmgfor Causationr"，Economic Inquiry，Vol. 17，pp. 401 ~413，1979.

⑪ E. L. Feige，and D. K. Pearce，"The Causal Relationship Between Money andIncome：Some Caveats for Time Series A-nalysis"，Journal of Econometrics，Vol. 39，pp. 199 ~211，1988.

⑫ C. W. J. Granger，"Some Recent Developments in a Concept of Causality"，Jour – nal of Econometries，Vol. 39，pp. 199 ~211，1988.

⑬ J. W. Pratt and R. Schlaifer，"On the Interpretation and Observation ofLaws"，Journal of Econometrics，Vol. 39，pp. 23 ~52，1988.

⑭ R. S. Pindyck，and J. J. Rotenberg，" Energy Shocks and theMacroeconomy"，in A. Alm and Weimer，Managing oil Shocks，Cambridge：Ballinger，1984.

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

设。参见 J. D. Hamilton 的“二战后的石油和宏观经济”[15]。这里我们转述其关于石油价格变化 ΔP_t 和真实 GNP 百分比变化 $\text{Log}(\text{GNP}_t / \text{GNP}_{t-1})$ 的因果关系检验。

Hamilton 进行 OLS 回归：

$$Z_t = a_0 + a_1 Z_{t-1} + \cdots + a_m Z_{t-m} + b_1 X_{t-1} + \cdots + b_m X_{t-m} + \varepsilon_t$$

首先令 $Z_t = \Delta P_t$，$X_t = (\text{GNP}_t / \text{GNP}_{t-1})$。然后，再反过来。下列结果来自两个根据 1973 年前季度数据的样本，统计量代表对 $b_1 = b_2 = \cdots = b_m = 0$ 的 F 检验，第一个 $m = 4$，然后 $m = 8$。

	$m = 4, N = 95$ (1942. 2—1972. 4)		$m = 8, N = 91$ (1950. 2—1974. 4)	
零假设	$F(4, 86)$	P 值	$F(8, 74)$	P 值
H_1：GNP → Oil	0. 58	0. 68	0. 71	0. 68
H_2：Oil → GNP	5. 55	0. 005	3. 25	0. 003

假设 H_2：石油价格变化不影响真实 GNP 变化在两种情况中都被严格拒绝，然而**假设 H_1：真实 GNP 的变化不影响石油价格变化**不能拒绝。这些结果和 Hamilton 提供的其它结果，证明了石油价格和美国经济之间有着十分密切的联系。

第二个例子来考察：先有鸡还是先有蛋?

这是自古以来就困扰着人们的一个问题。Thurman 和 Fisher 用因果关系检验对此进行的最新研究，终于对此争论给出了明确答案。这一例子引自 W. N. Thurman 和 M. E. Fisher 的“鸡蛋和因果关系，求谁更先?”[16]

此项研究用了两个变量的年度数据：1980 年至 1983 年的美国鸡蛋总产量（EGGS）和同期鸡的总产量（CHICKENS）。检验很简单，根据滞后的 EGGS 和 CHICKENS 对 EGGS 进行回归，如果滞后的 CHICKENS 的一组系数是显著的，那么“鸡生蛋”。然后，他们用一个对称的回归检验是否“蛋生鸡”。为得出两者中谁更先有的结论，必须找出单向性的因果关系。也就是说，拒绝“某一个不是另一个的原因”，同时，不能拒绝另一个不是某一个的原因。

---

⑮ J. D. Hamilton, "Oil and the Macroeconomy since World War Ⅱ", Journal of Political Economy, Vol. 91, pp. 228 ~ 248, 1983.

⑯ W. N. Thurman and M. E, Fisher, "Chickens, Eggs, and Causality, or Which Came First?", American Journal of Agriculture Economics, pp. 237 ~ 238, May. 1988.

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

Thurman 和 Fisher 的检验结果极富戏剧性，他们用 1 至 4 年的滞后值，断然拒绝了假设“蛋不影响鸡”；但却无法拒绝“鸡不影响蛋”这一假设。因此，他们得出结论，先有蛋！

Thurman 和 Fisher 指出，这种方法也可应用于其它重要问题的检验。例如，因果关系假设可能检验“谁最后笑，谁笑得最好”和“骄傲意味着毁灭，轻敌意味着失败”是否真实。

第七节　复习与提高

这一章反反复复地讲一个 F - 检验，并为此做了许多准备工作。t - 检验一般仅能够对含有单个参数的假设进行检验。对稍复杂点的单等式的假设，引入一个变量后，可以用 t - 检验进行检验。但如果假设中约束个数增多，就只好让 F - 检验帮忙了。反正无论是对单个参数还是多个参数或等式，F - 检验一概通用。

F - 检验的思路也很简单：

首先对原方程进行回归，得到一个未受约束的回归线差平方和 $ESS(u)$ 。

然后，我们不妨承认假设条件全部正确，并将之代入到原方程中合并重组后，构造出一个新方程，对新方程进行回归，得到一个新的受约束的残差平方和 $ESS(R)$ 。

再次，计算出统计量 F_0 : $F_0 = \dfrac{[ESS(R) - ESS(u)]/r}{ESS(u)/n-k-1}$ 。

最后，对 F_0 进行判断假设成立与否。

在 F - 检验的四种举例中，我们已不厌其烦地一次次温习 F - 检验的思路。这里不再重复了。讨论到自由度，又有瑟耳（Theil）教授发明的成果：带自由度的 $\bar{R}^2$ 。$\bar{R}^2$ 让我们开了一次眼界，如果你对它还有模糊不清的认识，想想那两幅有关 R^2 和 $\bar{R}^2$ 的图就全明白了。

图 7.3　描述了 F - 检验的思路。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



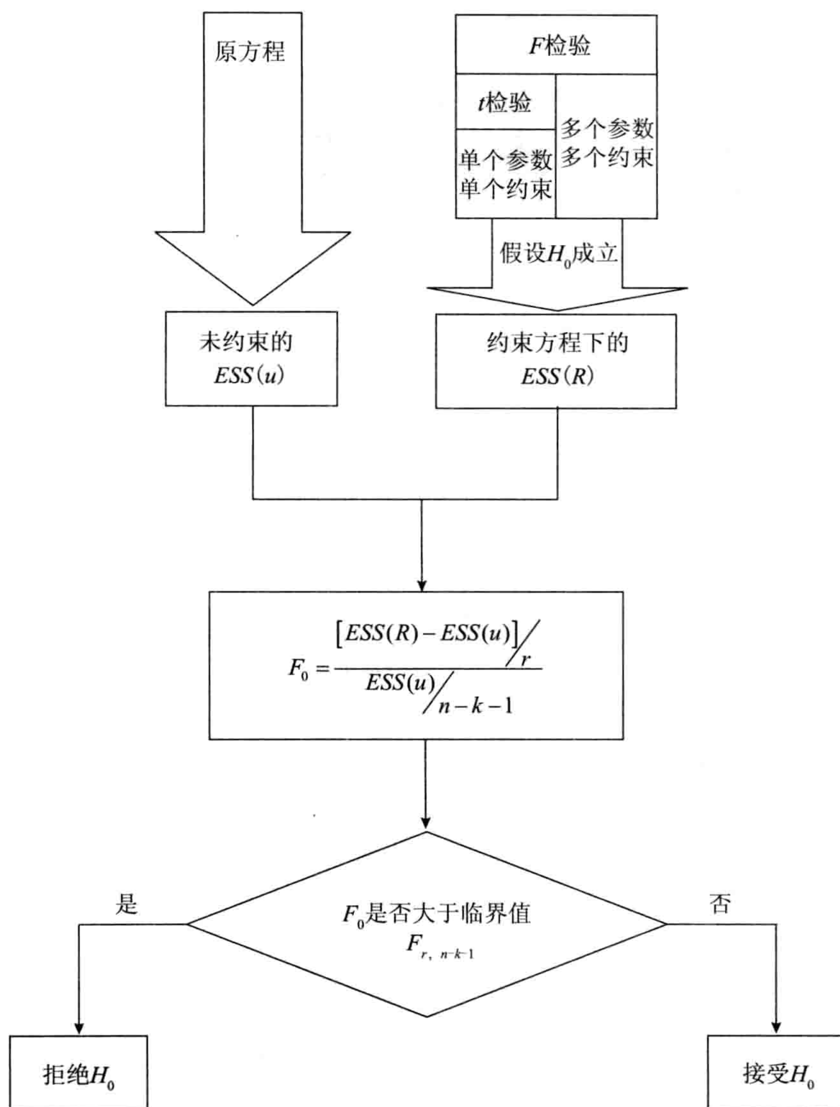


图 7.3

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第八章 虚拟变量

虚拟变量（Dummy Variable），又译虚设变量、名义变量、或哑变量，指的是一种取值为 0 或 1 的变量。

引入虚拟变量将使得线性回归模型变得更为复杂，也更接近现实。比如研究性别与收入（y_i）的关系：我们可以定义虚拟变量 D_i，$D_i=0$ 时表示女性，$D_i=1$ 时表示男性，即 $D_i=\begin{cases}1(男)\\0(女)\end{cases}$。对于线性回归模型 $y_i=\alpha+\beta D_i+\varepsilon_i$ 而言，若假设 $H_0:\beta=0$ 成立，则说明收入与性别将没有太大关系；如果假设 $H_0:\beta=0$ 不成立，则说明收入与性别有关。在 $\beta\neq0$ 时，借助 D_i，y_i 可分解成两个式子：$y_i=\begin{cases}\alpha+\beta(男)\\\alpha(女)\end{cases}$。

再看一个例子，研究就业与失业之间的收入（y_i）差异。定义虚拟变量 D_i，$D_i=1$ 表示就业，$D_i=0$ 表示失业。如果用线性回归模型 $y_i=\alpha+\beta D_i+\varepsilon_i$ 研究就业与失业的收入差异，那么，在就业时，收入 y_i 为 $y_i=\alpha+\beta$；失业时，收入 y_i 为 $y_i=\alpha$。

可见，D_i 是一个解释变量，它把可能的两种情况定义为 0 或 1，从而使一个方程能达到两个方程的作用。

D_i，能否用作为被解释变量呢？当然可以。比如财政政策是否应变动，结果只有两个：变动与不变动。我们就可以用一个虚拟变量来表示，但它是因变量。又如是否需要对某项工程拨财政资金，结果只有两个：拨付与不拨付。不妨定义 $D_i=\begin{cases}1(拨付)\\0(不拨付)\end{cases}$。

以 C_i 表示学校教育程度，R_i 表示年龄，我们就可以对以下线性回归模型进行分析：$D_i=\alpha+\beta_1C_i+\beta_2R_i+\varepsilon_i$。

可见，虚拟变量可以作为解释变量也可以作为因变量使用。作为因变量时，不论 D_i 的方程式多么复杂，x_i 有何值，D_i 仅有两种可能结果：要么是 1，要么是 0。我们可以用图 8.1 来表示。在这种情况下，我们不能简单地用最小二乘法进行参数估计，也就是说，当

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



虚拟变量 D_i 作为因变量时不能直接地使用最小二乘法进行估计。

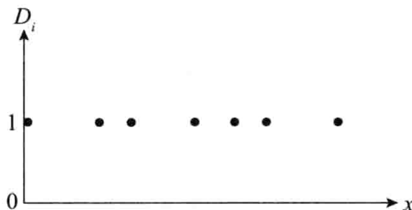


图 8.1

D_i 作为自变量时，也有一些隐患。看下面例子：

$$y_i = \alpha + \beta_1 D_{1i} + \beta_2 D_{2i} + \varepsilon_i$$

$$D_{1i} = \begin{cases} 0(\text{女}) \\ 1(\text{男}) \end{cases}$$

$$D_{2i} = \begin{cases} 0(\text{其他}) \\ 1(\text{怀孕者}) \end{cases}$$

在这种情况下，由于 $D_{1i} + D_{2i} = 1$，是典型的多重共线性问题。因此也不能简单地使用最小二乘法。

在定义 D_i 时，还要注意全集的范围，比如：已知全集为中、美、日三国国家政府。令

$$D_{1i} = \begin{cases} 1(\text{美}) \\ 0(\text{其他}) \end{cases}$$

$$D_{2i} = \begin{cases} 1(\text{中}) \\ 0(\text{其他}) \end{cases}$$

那么，在 $D_{1i} = D_{2i} = 0$ 时，就表示日本政府了。

$D_{1i} = 1$，$D_{2i} = 0$ 时，表示美国政府；

$D_{1i} = 0$，$D_{2i} = 1$ 时，表示中国政府。

因此根本不必要再定义一个 $D_{3i} = \begin{cases} 1(\text{日}) \\ 0(\text{其他}) \end{cases}$ 否则又会出现多重共线性问题，这个“其他”含义无穷，看一看图 8.2，就会更清楚一些。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

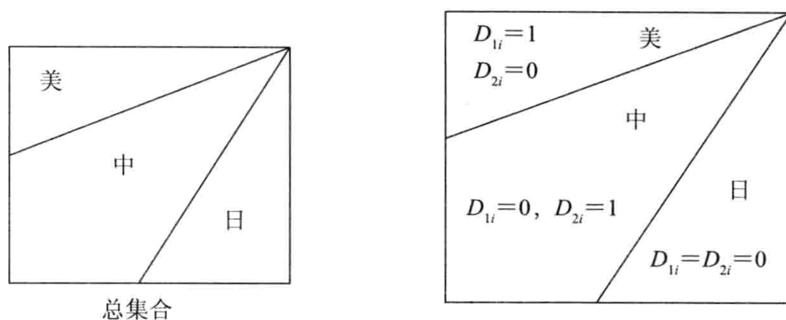


图 8.2

第一节　运用虚拟变量改变回归直线的截距

图 8.3 表示两种情况下，中国税收的变化情况：较低的直线描述了一般情况下税收收入与国民收入之间的关系；较高的直线则描述了 1994 年税制改革这一特殊年份后的变化情况。

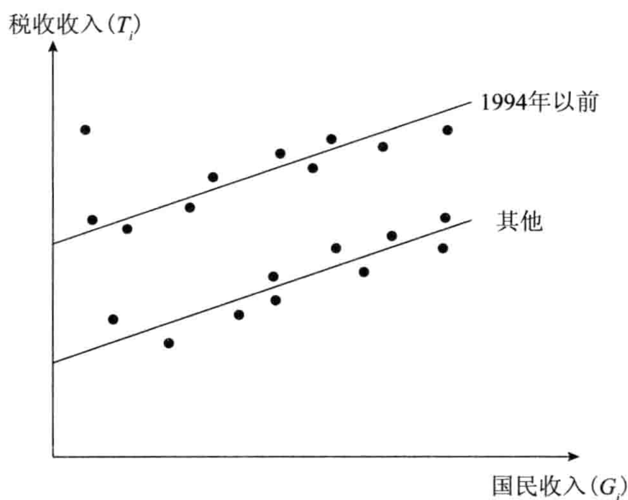


图 8.3

从这两条直线形状来看，它们趋势都相同，即这是两条平行的，但截距不同的两条直线，截距不同是因为预期基点不同。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



定义 $D_i = \begin{cases} 1 & (t \geqslant 1994) \\ 0 & (\text{其他}) \end{cases}$，建立线性回归模型：$T_i = \alpha + \beta_1 D_i + \beta_2 G_i + \varepsilon_i$，这样，我们可以用上面这个方程表示以下两种情况：

$$T_i = \begin{cases} (\alpha + \beta_1) + \beta_2 G_i + \varepsilon_i & (1994\text{ 年以后}) \\ \alpha + \beta_2 G_i + \varepsilon_i & (\text{其他}) \end{cases}$$

从上式我们可以看到1994年以后直线的截距为 $\alpha + \beta_1$，其他年份时，直线的截距为 α，因此，用一个方程就可以表示截距不同的两条直线。

第二节　运用虚拟变量改变回归直线的斜率

仍然研究税收收入和国民收入之间的相互关系。这一回假设，1994年以后与其他年份的预期基点相同，即截距相同，但变化幅度不同，也就是斜率不同。这是很常见的。税收收入在某一特定年份后异乎寻常地高涨，剧烈上升，与普遍年份的变化趋势大不相同，请参见图8.4。

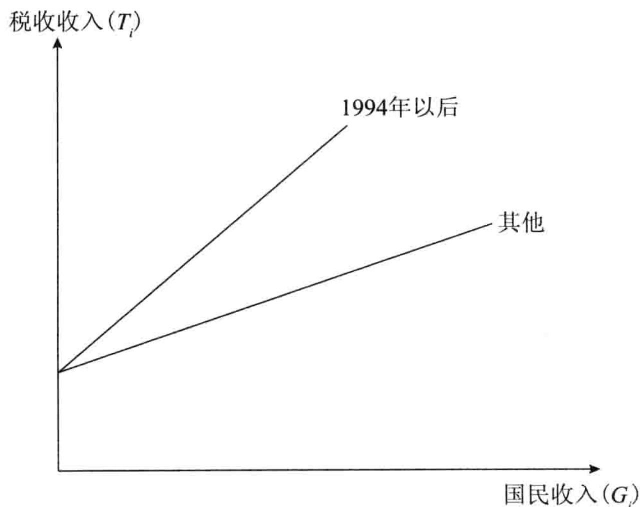


图8.4

仍定义：$D_i = \begin{cases} 1 & (t \geqslant 1994) \\ 0 & (\text{其他}) \end{cases}$。这次所要使用的线性回归模型要复杂些了，其形

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

式如下：

$$T_i = \alpha + \beta_1 G_i + \beta_2 (D_i \cdot G_i) + \varepsilon_i$$

这样，我们可以用上面的公式表示以下两种情况：

$$T_i = \begin{cases} \alpha + (\beta_1 + \beta_2) G_i + \varepsilon_i & (D_i = 1) \\ \alpha + \beta_1 G_i + \varepsilon_i & (D_i = 0) \end{cases}$$

由上式可见，直线的斜率变了，1994 年后直线的斜率是 $\beta_1 + \beta_2$，其余年份则是 β_1 。所以，我们同样运用含有虚拟变量的一个方程就表示了斜率不同的两种情况。

第三节　运用虚拟变量同时改变回归直线的斜率和截距

实际生活中，1994 年以后税收收入的变化与其他年份的根本不同。这种不同表现在税收收入的基点和趋势都不同。看一看图 8.5 你会觉得这种情况更符合实际。

要用一个方程表示出这样一种复杂的情形，就可以通过改变斜率、截距的复合形式构造出一个新方程来。仍定义：$D_i = \begin{cases} 1 & (t \geqslant 1994) \\ 0 & (\text{其他}) \end{cases}$ 。

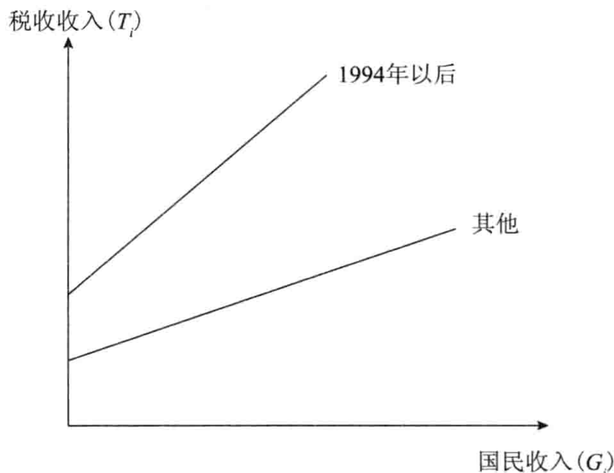


图 8.5

所构造的新方程具体形式如下：$T_i = \alpha + \beta_1 D_i + \beta_2 G_i + \beta_3 \cdot D_i \cdot G_i + \varepsilon_i$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



上面这个方程可以表示如下两种情况：

$$I_i = \begin{cases} (\alpha + \beta_1) + (\beta_2 + \beta_3) G_i + \varepsilon_i & (1994\text{ 年以后}) \\ \alpha + \beta_2 G_i + \varepsilon_i & (\text{其他}) \end{cases}$$

这样一来，回归直线的斜率和截距都发生了变化。

到了这里，也许会有小小的疑问：为什么非要用虚拟变量，而不直接对下面的两个方程：

$$T_i = (\alpha + \beta_1) + (\beta_2 + \beta_3) G_i + \varepsilon_i$$

$$\text{和 } T_i = \alpha + \beta_2 G_i + \eta_i$$

进行回归呢？它与对 $T_i = \alpha + \beta_1 D_i + \beta_2 G_i + \beta_3 \cdot D_i \cdot G_i + \varepsilon_i$ 进行回归之后再写成分裂式有什么区别呢？其区别在于随机扰动项 ε_i。如果直接对 $T_i = (\alpha + \beta_1) + (\beta_2 + \beta_3) G_i + \varepsilon_i$ 和 $T_i = \alpha + \beta_2 G_i + \eta_i$ 进行回归，这是两个单独方程，所以相当于进行了两次回归。现在的问题是我们不能保证两次回归中两个随机扰动项是同样的，这就意味着，我们是在不同的随机影响下分别研究 1994 年以后的情况和其他年份情况。而 $T_i = \alpha + \beta_1 D_i + \beta_2 G_i + \beta_3 \cdot D_i \cdot G_i + \varepsilon_i$ 则仅需一次回归，ε_i 是相同的，意味着无论是税制改革以后还是其他时间，随机因素的影响都假设是一样的。

第四节　折线回归

借助虚拟变量，我们已经能解决实际中一些常见的问题。但是在实际中，很可能会遇到折线情况，如图 8.6 所示：

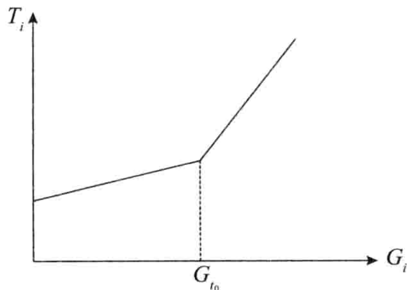


图 8.6

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

该图表示，在国民收入达到一定水平后，税收收入会大幅度上升。我们必须承认这个世界无奇不有，较高的经济增长率，可能会引起税收收入超常增长。因此，对图中所表示的情况也就不难理解了。

不妨假设，这个 t_0 为 1994 年，G_{t0} 为 1994 年的国民收入

$$D_i = \begin{cases} 1 & (t \geq 1994) \\ 0 & (\text{其他}) \end{cases}$$

我们可以使用下面这个方程对图中所示折线进行回归分析：

$$T_i = \alpha + \beta_1 G_i + \beta_2 (G_i - G_{t0}) D_i + \varepsilon_i$$

上述方程表示以下两种情况：

$$T_i = \begin{cases} (\alpha - \beta_2 G_{t0}) + (\beta_1 + \beta_2) G_i + \varepsilon_i & (t \geq 1994) \\ \alpha + \beta_1 G_i + \varepsilon_i & (\text{其他}) \end{cases}$$

依然得到一个截距和斜率都变化的方程，但这个方程的两个分支点是有接点的，如图 8.6 所示，当 $G_i = G_{t0}$ 时，两个分支处的值相等，因此两个方程相互连接在一起。

思考一下，如何用含有虚拟变量的方程估计图 8.7 所表示的情况呢？

对于这个问题，我们可定义以下两个虚拟变量：

$$D_{1i} = \begin{cases} 1 & (t \geq 1985) \\ 0 & (\text{其他}) \end{cases} \qquad D_{2i} = \begin{cases} 1 & (t \geq 1994) \\ 0 & (\text{其他}) \end{cases}$$

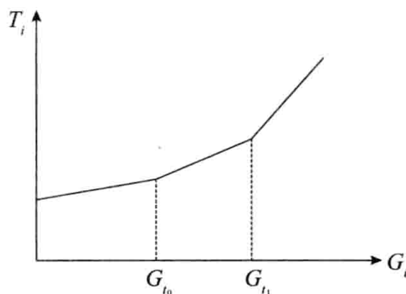


图 8.7

那么，我们就可以用下面的方程表示图中情况：

$$I_i = \alpha + \beta_1 G_i + \beta_2 (G_i - G_{t_0}) D_{1i} + \beta_3 (G_i - G_{t_1}) D_{2i} + \varepsilon_i$$

上述方程可以分解成：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



$$I_i = \begin{cases} \alpha + \beta_1 G_i + \varepsilon_i & (D_{1i} = D_{2i} = 0, 1985\text{ 年之前}) \\ (\alpha - \beta_2 G_{t_0}) + (\beta_1 + \beta_2) G_i + \varepsilon_i & (D_{1i} = 1, D_{2i} = 0, 1985\text{—}1994\text{ 年期间}) \\ (\alpha - \beta_2 G_{t_0} - \beta_3 G_{t_1}) + (\beta_1 + \beta_2 + \beta_3) G_i + \varepsilon_i & (D_{1i} = D_{2i} = 1, 1994\text{ 年之后}) \end{cases}$$

第五节　复习与提高

一、有关虚拟变量的复习与提高

虚拟变量的作用已令我们越来越感到吃惊。尽管它仅有 0 或 1 两个值，但却能不断地归拟出各种复杂的图形，或将多个方程归拟于一个方程中。

在定义虚拟变量时，要注意虚拟变量的范围问题。尤其是定义多个虚拟变量时，要格外注意每个虚拟变量的范围界限。以图 8.7 为例，如果定义的 D_{1i}、D_{2i} 分别为：

$$D_{1i} = \begin{cases} 1 & (t \geqslant 1985) \\ 0 & (\text{其他年份}) \end{cases}$$

$$D_{2i} = \begin{cases} 1 & (t \geqslant 1994) \\ 0 & (\text{其他年份}) \end{cases}$$

会怎样呢？将会有重合的 1985 ~ 1994 年这一段图形（见图 8.8）。

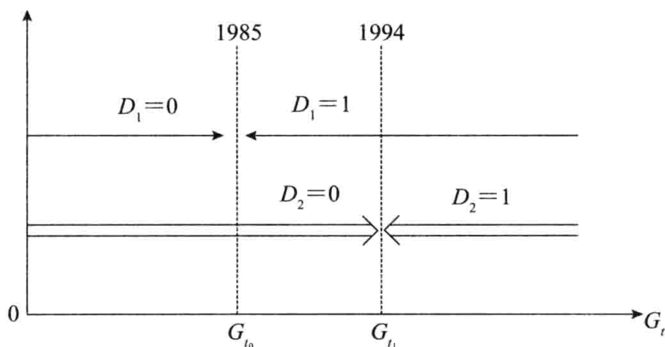


图 8.8

→和 ⇒ 分别指 D_{1i}、D_{2i} 中“其他年份”的范围，当 $D_{1i} = D_{2i} = 0$ 时，得到的不仅是 $(0, G_{t_0})$ 这一区域，而且还包括 (G_{t_0}, G_{t_1}) 和 $(G_{t_1}, +\infty)$。

这是违反我们本意的，我们原本要分段限制，所以这里必须使用分段的标号：≥、

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

≤；并以 t_0，t_1 为分界点。

当然，若定义 $D_{1i}=\begin{cases}1(1985\sim1994)\\0(\text{其它年份})\end{cases}$ 也不能成立，因为这样实际是说，在 1985 ~ 1994 年范围内 $D_{1i}=1$，小于 1984 年或大于 1994 年都将为 0，这样定义的虚拟变量代回我们使用的回归方程中去，将不会出现接点。因此，也不符合我们的初衷。

我们可以借助图 8.9 看到正确定义的高明之处。

$$D_{1i}=\begin{cases}1(t\geq1985)\\0(\text{其他})\end{cases}$$

$$D_{2i}=\begin{cases}1(t\geq1994)\\0(\text{其他})\end{cases}$$

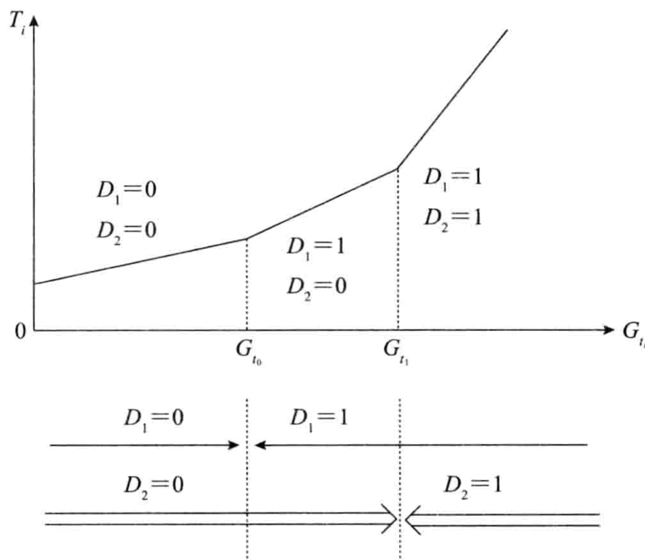


图 8.9

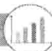
这才符合我们的本意。所以在定义虚拟变量时，千万要小心这个“其他”的实际范围。

以前我们只解决了两段折线和三段折线情况，如果有四段、五段、六段……折线，就是有三个折点、四个折点、五个折点……的情况下怎么办呢？容易得很：

有几个折点，就定义几个虚拟变量，即在方程中增加几个 β_i 即可。

可比下，找出规律来，你会发现这一切都简单得很，参见表 8 - 1。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



定义虚拟变量时，要以这些折点为分界点，假如第 $i+1$ 个折点对应 G_{t1} ，那么，我们可以定义：

$$D_i = \begin{cases} 1 & t \geq t_i \\ 0 & \text{其他} \end{cases}$$

表 8－1

虚拟变量	折线点个数	图形	公式
$D_i = \begin{cases} 1 & t \geq t_0 \\ 0 & （其他） \end{cases}$	1	G_{t_0}	$T_i = \alpha + \beta_1 G_i + \beta_2 (G_i - G_{t_0}) D_i + \varepsilon_i$
$D_{1i} = \begin{cases} 1 & t \geq t_0 \\ 0 & （其他） \end{cases}$ $D_{2i} = \begin{cases} 1 & t \geq t_1 \\ 0 & （其他） \end{cases}$	2	G_{t_0} G_{t_1}	$T_i = \alpha + \beta_1 G_i + \beta_2 (G_i - G_{t_0}) D_{1i} + \beta_3 (G_i - G_{t_1}) D_{2i} + \varepsilon_i$
$D_{1i} = \begin{cases} 1 & t \geq t_0 \\ 0 & （其他） \end{cases}$ $D_{2i} = \begin{cases} 1 & t \geq t_1 \\ 0 & （其他） \end{cases}$ $D_{3i} = \begin{cases} 1 & t \geq t_2 \\ 0 & （其他） \end{cases}$	3	G_{t_0} G_{t_1} G_{t_2}	$T_i = \alpha + \beta_1 G_i + \beta_2 (G_i - G_{t_0}) D_{1i} + \beta_3 (G_i - G_{t1}) D_{2i} + \beta_4 (G_i - G_{t2}) D_{3i} + \varepsilon_i$

2. 虚拟变量的通观

通过上面的讨论，我们可以看出，根据图形的不同，我们可以采取不同的方程形式进行描述。这一章主要讨论了表示以下四种图形的四种方程（见表 8－2）。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

表 8－2

	图形	相对应的方程
准备阶段	T_i, 0, G_i; T_i, G_i	$T_i = \alpha + \beta_1 D_i + \beta_2 G_i + \varepsilon_i$； $T_i = \alpha + \beta_1 D_i + \beta_2 (D_i G_i) + \varepsilon_i$
第一类复合	T_i	$T_i = \alpha + \beta_1 D_i + \beta_2 G_i + \beta_3 D_i G_i + \varepsilon_i$
第二类复合	T_i, G_{t_0}, G_i	$T_i = \alpha + \beta_1 G_i + \beta_2 (G_i - G_{t_0}) D_i + \varepsilon_i$ $D_i = \begin{cases} 1(t \geqslant t_0) \\ 0(\text{其他年份}) \end{cases}$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第九章 计量经济学软件 EViews 简介

EViews 是在大型计算机的 TSP（Time Series Processor）软件包基础上发展起来的新版本，是处理时间序列数据的有效工具。1981 年 Micro TSP 面世，1994 年 QMS（Quantitative Micro Software）公司在 Micro TSP 基础上直接开发成功 EViews 并投入使用。EViews 能为我们提供基于 Windows 平台的复杂的数据分析、回归及预测工具，通过 EViews 能够快速从数据中得到统计关系，并根据这些统计关系进行预测。EViews 在系统数据分析和评价、金融分析、宏观经济预测、模拟、销售预测及成本分析等领域中有着广泛的应用。

目前，已有多种统计软件包系统，各有特色。如著名的 SAS 系统，全称是 Statistic Analysis System（统计分析软件包），适宜在 IBM 大型机上进行统计分析，并专门成立有 SAS 公司，从事 SAS 系统软件的开发。目前已发展成为一个十分庞大的系统，其中 Micro－SAS 为 SAS 的子系统，它是一个能够在微机上运行的统计软件包。除此之外，SPSS 也是著名的统计处理软件包，开发比 SAS 要稍晚一些，其全称是 Statistic Package for Social Science（社会科学统计软件包）。世界上著名的计算机公司，如 IBM 公司、DEC 公司和 HP（惠普）公司等机器上都配备了上述二种统计软件包或其它统计软件包，这里就不一一评述了，我们把注意力集中到 EViews 上来。

第一节 关于计算机的一些基本概念

一、计算机系统简介

计算机是由硬件和软件两个部分组成的。

硬件 外观上看，一台计算机由主机、显示器、键盘、鼠标器和音箱等部件所组成。功能上看，计算机的硬件主要包括中央处理器、存储器、输入设备、输出设备等。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

中央处理器 中央处理器（CPU）是计算机的“心脏”，它安装在主机箱内。计算机中的一切工作都通过它来处理。中央处理器能进行复杂的运算，控制各个设备协调一致地工作。

存储器 是计算机系统中的记忆设备，用来存放程序和数据。计算机中的全部信息，包括输入的原始数据、计算机程序、中间运行结果和最终运行结果都保存在存储器中。它根据控制器指定的位置存入和取出信息。分为内存储器、外存储器。

输入设备 是计算机用来接受指令和数据等信息的。常用的输入设备有键盘、鼠标器等。

输出设备 是计算机负责传送处理结果的设备。常用的输出设备有显示器、打印机、音箱等。

软件 分为系统软件和应用软件两大类。

系统软件 是一种管理计算机硬件和为应用软件提供运行环境的软件，如 UNIX、Windows、Linux 等都是系统软件。

应用软件 是为了完成某种特定用途而编制的软件。有了应用软件，才能在计算机上画图、写文章，制作多媒体报告、玩游戏等，如 Word、PowerPoint 以及我们即将要学习的 Eviews 等都是应用软件。

二、数据处理的一般过程

数据处理的一般过程可以划分为三大步：首先，我们必须把数据和程序放到计算机中去，才有可能让计算机进行处理，这一过程为输入过程（INPUT）；其次，对数据进行加工处理，这个过程叫处理过程（PROCESS）；最后，在数据经过处理后还要把处理结果输出出来，这个过程叫输出过程（OUTPUT）。

根据以上思路，研究一下三大步骤的具体形式：

输入过程 分为两种：一种是从纸上输入，一种是从磁盘上输入。

处理过程 在进行数据处理时，面临的问题是：（1）如何把数据放到计算机中去，这是个数据结构问题（Data Structure）；（2）对数据进行加工的命令有哪些，这是处理命令的问题（Process Command）。处理命令可分为两类：一般命令（General Command）和特殊功能命令（Special Function）。所谓一般命令，是指各类计算机语言或软件包都必须具备的一些基本功能，如处理语句、条件语句和循环语句。而特殊功能命令，则是不同语言或软件包所具有的不同的特定功能。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



输出过程 有三种形式：纸上输出、磁盘输出和屏幕输出。所谓输出，实际上就是指在哪儿把处理结果表示出来的问题。

三、计算机语言的一般结构

目前有多种计算机语言流行，如 BASIC、Pascal、C、Java、C++等，但是它们都有一些类似或者相同的一般命令。根据我们对数据处理过程的认识，可以归纳出任何计算机语言或软件包的一般结构。任何一门计算机语言，都应包括以下三类命令：输入命令（INPUT COMMAND）、处理命令（PROCESSING COMMAND）和输出命令（OUTPUTCOMMAND）。以 BASIC 语言为例，它的输入命令有：Input、Read - Data、Load、Restore；它的输出命令有 Print、Save、Lprint；它的处理命令有：Let、If … Then … Else、For … Next、Gosub、Return、Goto 等等。除此之外，BASIC 的数据结构可分四类：（1）按常量存放；（2）按变量存放；（3）按数组变量存放，所谓数组就是指单下标、双下标的下标变量；（4）按文件存放。

第二节 EViews 概述

EViews 处理的主要对象是时间序列，每个序列都有一个名称，只要提出序列的名称就可以对序列中所有的数据进行操作。它允许用户从键盘，磁盘文件输入得到数据，并能从已有的数据得到新的数据，显示和打印数据，做数据序列的统计分析和相关分析。EViews 得益于 Windows 的可视的特点，能通过标准的 Windows 菜单和对话框，用鼠标选择操作，并且能通过标准的 Windows 技术来使用显示于窗口中的结果。

一、启动软件包

在 EViews 软件包安装完成后，桌面将生成图标“ ”。双击图标，或点击开始→程序→EViews，进入 EViews 的主界面窗口（见图 9.1）。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

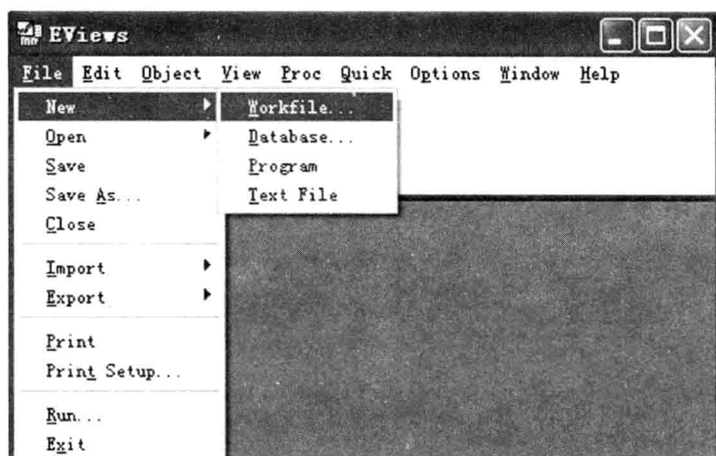


图 9.1

二、EViews 的窗口

EViews 的窗口分为标题栏、菜单栏、命令窗口、状态行和工作区。

标题栏：它位于主窗口的最上方。当 EViews 工作区窗口处于活动状态时，工作区窗口的标题栏的颜色较其他窗口比是蓝色的，当其他窗口处于活动状态时，它的颜色会变成灰色的。可以单击 EViews 工作区窗口的任何位置使 EViews 工作区窗口回到活动状态。

主菜单：紧接着标题栏下面是主菜单。移动光标至主菜单然后点击鼠标左按钮，它会出现一个下拉菜单，在这个下拉菜单中可以单击选择显现项。

命令窗口：菜单栏下面是命令窗口。把 EViews 命令输入该窗口，按回车键即执行该命令。该窗口支持 Windows 下的剪切和粘贴功能，因此可以在命令窗口、其他的 EViews 文本窗口及其它的 Windows 窗口之间转换文本。该命令窗口中的内容能被直接保存到一个文本文件中：通过单击窗口的任何位置确定命令窗口当前处于活动状态，然后从主菜单上选择 File/Save As。可把光标放在命令窗口的最底端，按着鼠标按钮上下拖拽来改变命令窗口的大小。

状态栏：窗口的最底端是状态栏，它被分成几个部分。左边部分有时提供 EViews 发送的状态信息，通过单击状态线最左边的方块可清除这些状态信息；往右接下来的部分是 EViews 寻找数据和程序的预设目录；最后两个部分显示预设数据库和工作文件的名称。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

工作区域：位于窗口中间部分的是工作区。EViews 在这里显示各个目标窗口，这些窗口会相互重叠且当前活动窗口处于最上方，只有活动窗口的标题栏是深色的。当需要的窗口被部分覆盖时，可单击该窗口的标题栏或该窗口的任何可见部分使该窗口处于最上方；此外，还可通过单击菜单、选择需要的窗口名称来直接选择窗口。移动窗口可通过单击标题栏并拖拽窗口来完成。单击窗口右端底部的角落并拖拽角落可改变窗口的大小（见图 9.2）。

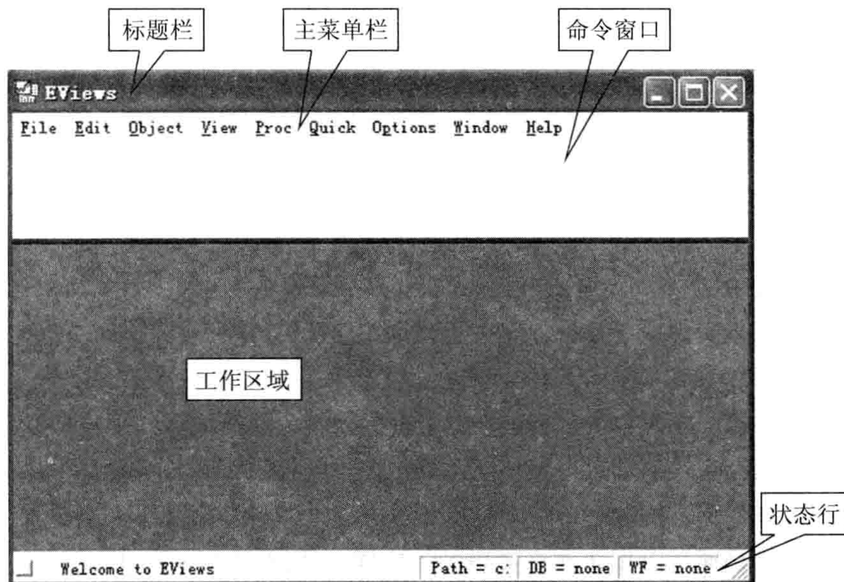


图 9.2

三、关闭 Eviews。

这里有关闭 EViews 的许多方法：

1. 可在主菜单上选择 File/Close 或按 ALT – F4 键来关闭 EViews。

2. 如果正在运行，可单击 EViews 窗口右上角的关闭方块，或双击 EViews 窗口左上角的 EVIEWS 符号来关闭窗口。

3. 单击 EViews 窗口左上角的控制菜单方块，然后选择 Close 来关闭窗口。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第三节 工作文件的建立及相关操作

EViews 的核心是对象，对象是指有一定关系的信息或算子捆绑在一起供使用的单元，用 EViews 工作就是使用不同的对象。对象都放置在对象集合中，其中工作文件（workfile）是最重要的对象集合。

一、什么是工作文件

工作文件是 EViews 对象的集合。EViews 中的大多数工作都涉及对象，它们包含在工作文件中，因此使用 EViews 工作的第一步是创建一个新的工作文件或调用一个已有的工作文件。

每个工作文件包括一个或多个工作文件页，每页都有它自己的对象。一个工作文件页可以被认为是子工作文件或子目录，这些子工作文件或子目录允许我们在工作文件内组织数据。工作文件可以容纳一系列 EViews 对象，如方程、图表和矩阵等。

因为工作文件页的主要目的是容纳数据集合的内容，因此每页必须包括观测值标识符的信息。一旦给出标识符的信息，工作文件页将在与此相联系的数据集合中提供与观测值相关的内容，允许我们应用数据、处理延迟、或者运用纵向的数据结构。

大多数工作都是通过工作文件来实现的。这样，使用 EViews 工作的第一步就是建立一个新的工作文件或调用一个已有的工作文件。

二、创建工作文件

（1）选择菜单 File→New→workfile，打开 Workfile Create 对话框，如图 9.3 所示。

对话框的左边 Workfile structure type 下方是下拉列表框，它用来描述数据集合的基本结构。可以在 Dated - regular frequency，Unstructured 和 Balanced Panel 中选择。一般来说，若是一个简单的时间序列数据集合，可以选择 Dated - regular frequency，对于一个简单的面板数据库，可以使用 Balanced Panel，而在所有其他情况下，可以选择 Unstructured。

当选择 Dated - regular frequency 时，EViews 将允许选择数据的频率。见右侧 Date specification Frequency 中可以在下面两者之间进行选择，一个是标准的 EViews 所支持的数据频率（Annual（年度）、Semi - annual（半年度）、Quarterly（季度）、Monthly（月

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

度）、Weekly（周度）、Daily－5 day week（每 5 天一个星期）、Daily－7 day week（每 7 天一个星期））；另外一个是特定的频率（Integer date）。选择频率时，要正确设置数据中观测值的间隔，（无论它们是年度，半年度，季度，月度，周度，每周 5 天，还是每周 7 天），以便于允许 EViews 使用所有可用的日历信息来组织和管理数据。随后为工作文件输入 Start date 和 End date。点击 OK，EViews 将创建一个具有固定频率的工作文件。

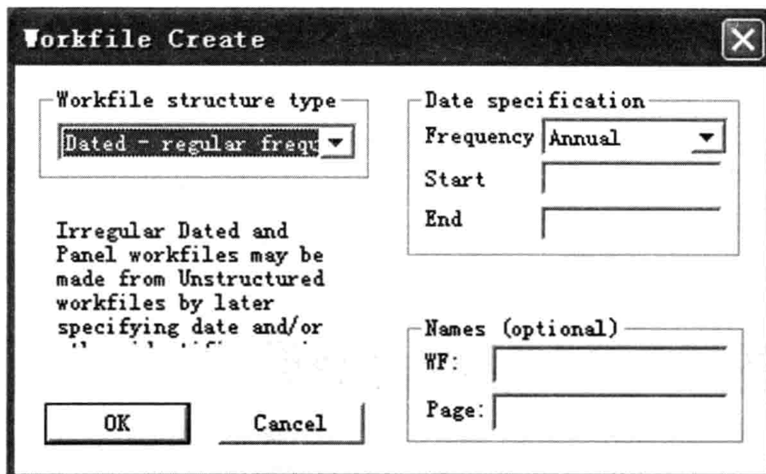


图 9.3

我们以我国国内生产总值（GDP）和税收收入总额（tax）为例，说明 EViews 的具体用法。其数据见表 9－1：

表 9－1

年份	GDP（亿元）	tax（亿元）
1989	16992	2727
1990	18668	2822
1991	21781	2990
1992	26924	3297
1993	35334	4255
1994	48198	5127
1995	60794	6038

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

续前表

年份	GDP（亿元）	tax（亿元）
1996	71177	6910
1997	78973	8234
1998	84402	9263
1999	89677	10683
2000	99215	12582
2001	109655	15301
2002	120333	17636
2003	135823	20017
2004	159878	24166
2005	183085	28779

我们新建工作文件（workfile），见图 9.4。

图 9.4

点击 OK 便得到工作文件如图 9.5。

由图 9.5 可知，有两个小图标β c和 resid，在此新建文件中，这两个图标分别代表了系数向量 C 和残差向量 RESID.

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



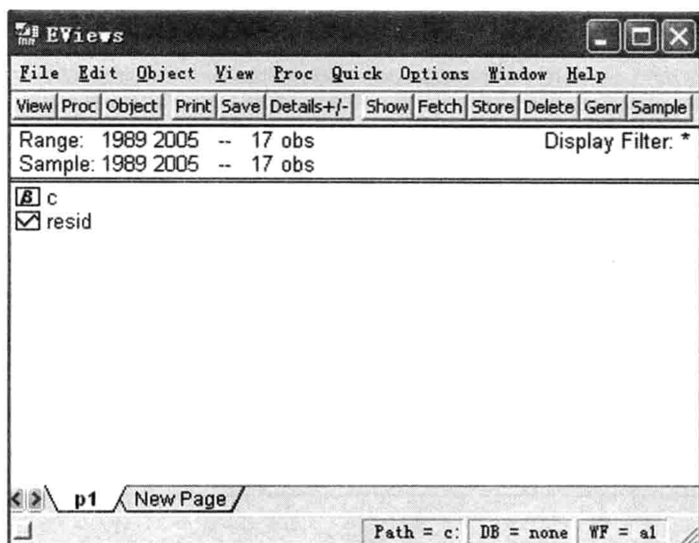


图 9.5

三、导入数据

点击 EViews 主界面中 File→Import→Read Test – Lotus – Excel，进入导入文件对话框口，见图 9.6：

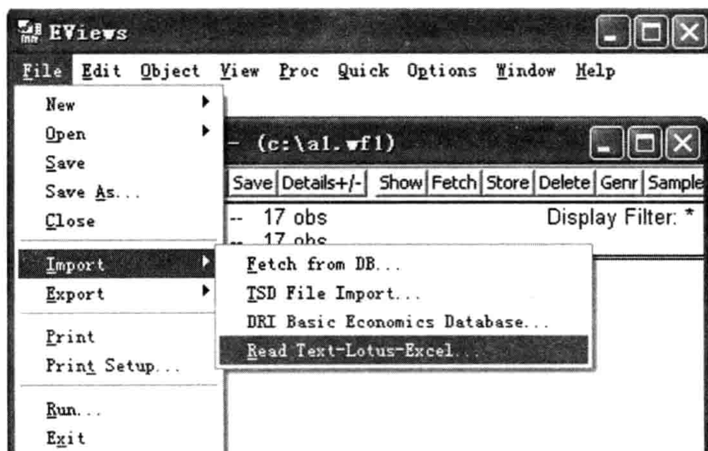


图 9.6

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

出现对话框如图 9.7：



图 9.7

选中 Excel 文件，出现对话框如图 9.8：

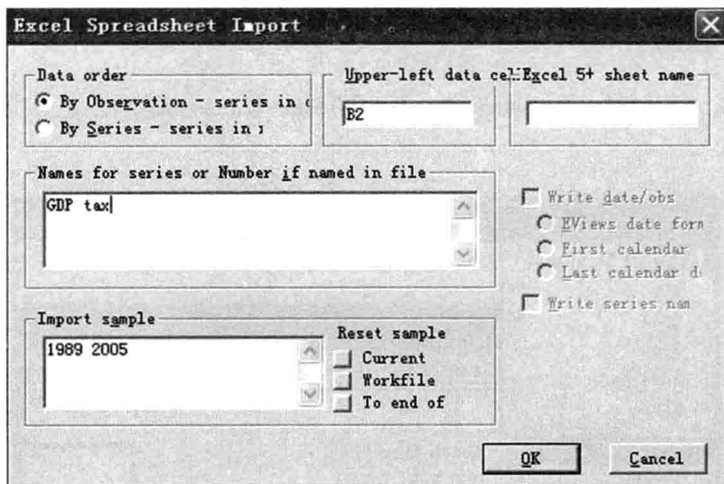


图 9.8

本例中 GDP（1989－2005）.xls 文件中有两个变量，将两个变量的名字 GDP 和 tax 填入 Name for series or Number if named in file 下面的框中，然后点击 OK，文件导入完成。

手工输入数据部分就不再详述，请参阅使用手册相关资料。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



四、数据的基本处理

EViews 提供了多种方法将数据进行处理。这里举一个例子，用 EViews 将刚刚导入的数据绘制成曲线。

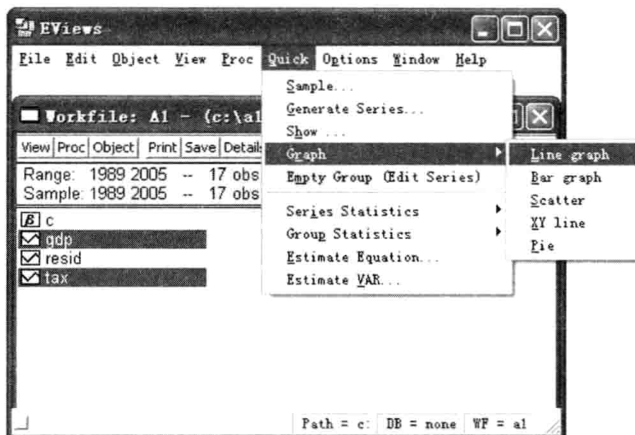


图 9.9

将 GDP 和 tax 项选中，点击菜单 Quick→Graph→Line graph 即可生成如图 9.10：

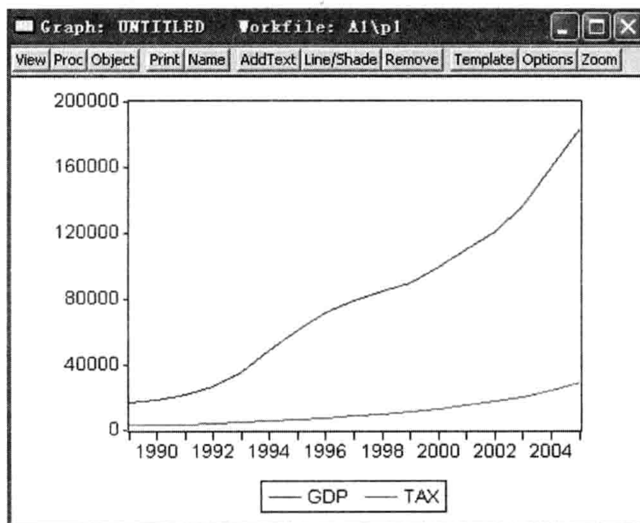


图 9.10

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第四节　基本回归模型

单方程回归是最丰富多彩和广泛使用的统计技术之一。本节介绍 EViews 中基本回归技术的使用，对于更进一步的应用如加权最小二乘法、二阶段最小二乘法（TSLS）、非线性最小二乘法、ARIMA/ARIMAX 模型、GMM（广义矩估计）、GARCH 模型和定性的有限因变量模型，这些技术和模型都是建立在本节介绍的基本思想的基础之上的，大家可以在掌握了基本回归模型的基础上，自行扩展学习。

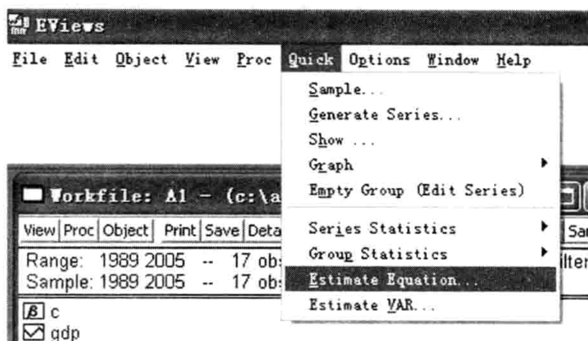


图 9.11

在 EViews 主菜单上选择 Quick→Estimate Equation，出现对话框，见图 9.11。

在图 9.11 Equation Specification 对话框中，先后输入被解释变量 tax，常数项 c 和解释变量 GDP，点击 ok，这样，统计的回归结果就出来了，见图 9.12。

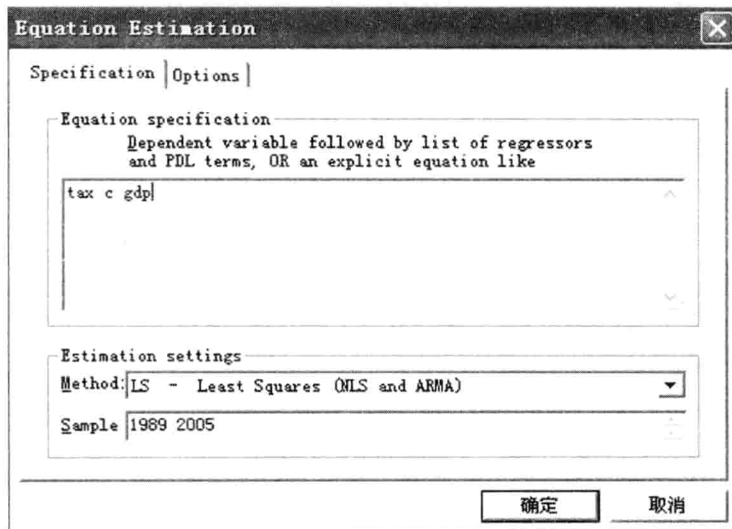


图 9.12

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

Equation: UNTITLED Workfile: A1\p1

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: TAX
Method: Least Squares
Date: 02/27/09 Time: 15:31
Sample: 1989 2005
Included observations: 17

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1860.686	734.6793	-2.532650	0.0230
GDP	0.156115	0.007836	19.92371	0.0000

R-squared	0.963588	Mean dependent var	10636.88
Adjusted R-squared	0.961161	S.D. dependent var	8001.807
S.E. of regression	1576.971	Akaike info criterion	17.67453
Sum squared resid	37302555	Schwarz criterion	17.77256
Log likelihood	-148.2335	F-statistic	396.9541
Durbin-Watson stat	0.182399	Prob(F-statistic)	0.000000

图 9.13

由结果，我们可以得到一元线性回归方程

$$\begin{aligned} y_i &= -1860.686 + 0.156115x_i \\ &\quad (734.6793) \qquad (0.007836) \end{aligned}$$

。

第五节 复习与提高

EViews 软件包会像你后来发现的一样，十分容易使用，并可以扩展你的回归范围。我们将在日后的学习中随着知识的增加而学到许多种新的功能命令。慢慢地，你会发现，EViews 可以实现在以后的每一章里你都会接触到的内容。本章中给出的只是些基础。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第十章 异方差

到目前为止，我们已比较熟悉古典假设前提下的线性回归模型，并对它的一些重要性质也已进行了较为深入的讨论。然而在现实生活中遇到的问题却往往比古典模型复杂得多，这就使得古典模型中一些问题暴露出来了。我们已知古典模型得以畅行无阻的保护盾是它的六个前提假设：

假设 1　残差纵向变动（Vertical Jumps）

假设 2　$E(\varepsilon_i)=0$

假设 3　$Var(\varepsilon_i)=\sigma^2$（同方差）

假设 4　$\text{cov}(\varepsilon_i,\varepsilon_j)=0(i\neq j)$

假设 5　$\text{cov}(x_i,\varepsilon_j)=0$（$x_i$ 和 ε_j 无线性相关）

假设 6　数据产生过程为线性过程：$y_i=\alpha+\beta_1x_{1i}+\beta_2x_{2i}+\cdots+\beta_kx_{ki}+\varepsilon_i$

一旦这六条假设的其中之一不成立，那么，在六条假设条件下辛辛苦苦得到的高斯-马尔科夫定理就将是废纸一张。而在现实中，对这六个假设中每一个的放松，都会产生更为复杂的问题，需要使用新的方法来解决。我们不妨逐一来仔细看这六条假设。

假设 1　残差纵向变动（Vertical Jumps）

这条假设的意义在于，自变量 x_i 是非随机变量，而且它的观测值也都是百分之百地正确无误，不存在观测误差。而仅在因变量 y_i 上存在着随机残差 ε_i 这样的变化，就是所谓的纵向变化。但在现实生活中，x_i 的观测值也存在着观测误差的问题。这就是说，x_i 也存在一个水平变动的横向误差，并且在此基础上，因变量 y_i 还有它自己的纵向误差。以一元线性回归模型为例，假设真正的数据产生过程为：$y_i=\alpha+\beta x_i+\varepsilon_i$，那么含有横向波动误差时，方程则为：$y_i=\alpha+\beta(x_i+\eta_i)+\varepsilon_i$。

上面方程说明自变量是随机变量。因此，这种模型，就叫误差变量模型（Errors in Variables Model）。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

假设 2 $E(\varepsilon_i) = 0$

在这六个假设中唯有这一假设，在任何情况下都可以得到保证。仍以一元线性回归模型为例，在 $E(\varepsilon_i) = 0$ 下，肯定了 $E(y_i) = E(\alpha + \beta x_i + \varepsilon_i) = \alpha + \beta x_i$。如果 $E(\varepsilon_i) = r(r \neq 0)$，就会有：

$$
\begin{aligned}
E(y_i) &= E(\alpha + \beta x_i + \varepsilon_i) \\
&= E(\alpha + \beta x_i) + E(\varepsilon_i) \\
&= (\alpha + r) + \beta x_i
\end{aligned}
$$

令 $\alpha^* = \alpha + r$，则有：$E(y_i) = \alpha^* + \beta x$，因此，这并不会对回归估计量的性质产生严重影响，也就是说，我们可以把 $E(\varepsilon_i) \neq 0$ 时的期望值并入回归模型的常数项中，从而使假设得以保证。

假设 3 $Var(\varepsilon_i) = \sigma^2$

$Var(\varepsilon_i) = \sigma^2$ 称同方差，指 ε_i 的方差与其它任何变量之间无任何关系，$E(\varepsilon_i^2) = \sigma^2$ 为常量。如果 $Var(\varepsilon_i) = \sigma^2$ 不成立，换言之 $Var(\varepsilon_i) = \sigma_i^2$，就出现了**异方差**问题。它意味着随自变量的取值不同，ε_i 的方差就发生变化。

在现实中这种例子随处可见。例如探讨国民收入与财政支出问题，国民收入用 I_i 表示，财政支出用 E_i 表示。分别考察两个样本组，第一组省份在 3000 亿/年 ~4000 亿/年范围内，第二组在 5000 亿/年 ~1 万亿/年范围内，用同一个一元线性回归模型进行回归分析：

$$E_i = \alpha + \beta I_i + \varepsilon_i$$

进行两次回归分析后，你会感到所得到的残差是不同的。很显然，第一组收入水平很低，几乎需要全部用来维持行政管理费用与基础设施建设，这样它的回归残差会很小；第二组收入高的省份，在保证前述基本支出后会增加教育、科研、卫生等不同方面的支出，从而使回归残差变化较大。而同方差假设意味着不管国民收入是多少，每个省份支出行为一致。这当然与实际不符了。

假设 4 $\text{cov}(\varepsilon_i, \varepsilon_j) = 0 (i \neq j)$

这一假设认为 ε_i 与 ε_j 之间是毫无关系的，即使是 ε_i 与 ε_{i+1}，ε_{i-1} 与 ε_i 之间也如此。如果这一假设条件破坏了，即 $E(\varepsilon_i, \varepsilon_j) \neq 0$，就说明 ε_i 与 ε_j 之间存在相关关系，这就是所谓的**自相关**问题（Autocorrelation）。

假设 5 $\text{cov}(x_i, \varepsilon_j) = 0$（$x_i$ **和** ε_j **无线性相关**）

这条假设在自变量不止一个时，还应包括另外一个内容：$\text{cov}(x_i, x_j) = 0$。因此，这条

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

假设的完整内容应为：所有自变量与残差都不相关，$\text{cov}(x_i,\varepsilon_j)=0$，所有自变量之间也无关，$\text{cov}(x_i,x_j)=0$。如果这一假设中 $\text{cov}(x_i,x_j)=0$ 不成立，将产生**多重共线性**问题（Multicollinearity）。如果这一假设中 $\text{cov}(x_i,\varepsilon_j)=0$ 不成立，则会产生**联立方程组**问题。

假设 6　数据产生过程为线性的：$y_i=\alpha+\beta x_i+\varepsilon_i$

当非线性出现时，应设法将它变成线性。当然，并不是所有非线性方程都可以转化为线性方程，但绝大多数方程是可以线性化的。

例如 $y_i=\dfrac{1}{\alpha+\beta x_i}$，可转为 $\dfrac{1}{y_i}=\alpha+\beta x_i$。令 $y_i^*=\dfrac{1}{y_i}$，则 $y_i^*=\alpha+\beta x_i$。

在 Eviews 软件中，可用生成语句（GENERATE）来生成新的数据序列 $y_i^*=\dfrac{1}{y_i}$：

GENR　YR = 1/y

至此，我们已熟悉了在古典假设条件不成立时，可能产生的五个问题：误差变量模型、异方差、自相关、多重共线性和非线性问题。我们将把误差变量模型放到联立方程一章中讨论，而在第十章、第十一章中主要讨论有关异方差、自相关和多重共线性问题。

第一节　异方差概述

什么是异方差？它有哪些假设前提呢？在本章开始部分，我们已粗略了解到异方差与同方差的区别。在严格的定义下，异方差必须具有如下六个假设：

假设 10.1.1　残差纵向变动

假设 10.1.2　$E(\varepsilon_i)=0$

假设 10.1.3　$Var(\varepsilon_i)=\sigma_i^2$

假设 10.1.4　$\text{cov}(\varepsilon_i,\varepsilon_j)=0(i\neq j)$

假设 10.1.5　$\text{cov}(x_i,\varepsilon_j)=0$，$\text{cov}(x_{ij},x_{ej})=0$，$i\neq e$

假设 10.1.6　数据产生过程为线性：$y_i=\alpha+\beta_1x_{1i}+\beta_2x_{2i}+\cdots+\beta_kx_{ki}+\varepsilon_i$

换言之，古典假设中除假设 3 不成立外，即 ε_i 的方差 σ^2 换成 σ_i^2 外，异方差仍满足其余五条假设。值得注意的是，由于古典假设中的假设 3 已不成立，古典模型中高斯－马尔科夫定理将不再成立。

现在我们对异方差已有了一些了解，在本章中，将要对如下三个问题进行讨论，一是如何发现异方差；二是在存在异方差的条件下如果还用最小二乘法，那么最小二乘估计量

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

将产生什么后果；三是如何解决异方差问题。

第二节　如何发现异方差

必须承认，运用电子计算机能够给我们以很大帮助，EViews 将会给我们许多直观结果。这就是发现异方差的直观方法：图示法（Graphical Method）。所谓图示法，就是运用 EViews 中的“GRAPH”命令画出能够反映拟合残差值和自变量 x_i 之间相互关系的图来，即用回归得到的拟合残差值 $\hat{\varepsilon}_i$ 和 x_i 作图。一般说来 $\hat{\varepsilon}_i$ 有如下情况（见图 10.1）：

1. 表示 $|\varepsilon_i|$ 随 x_i 的增大而增大；
2. 表示 ε_i 随 x_i 的增大而减少；
3. 表示 ε_i 随 X_i的增加而增加。

下面的几种异方差表现的是一些更为复杂的情况，有的甚至可能牵涉到多元结构。

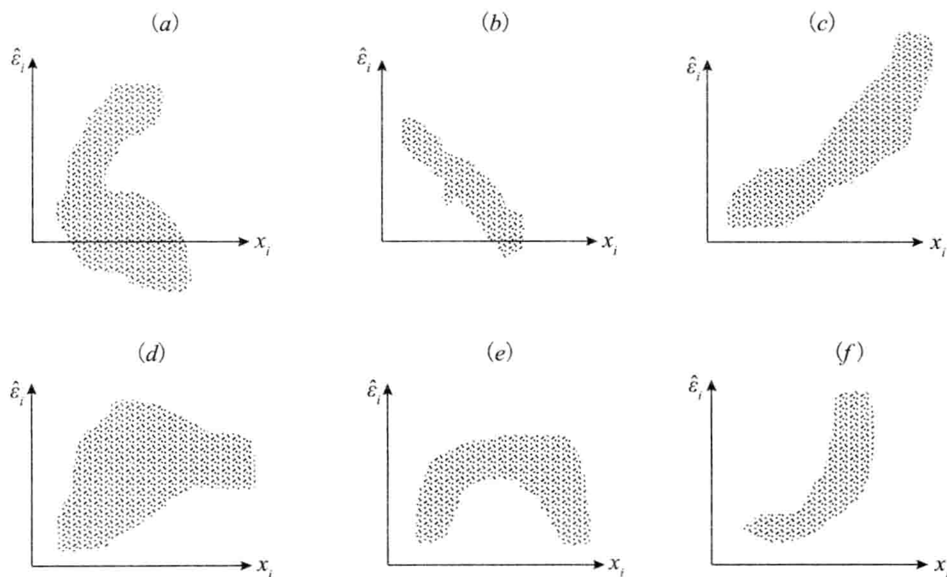


图 10.1

直观地说，图示法就是运用图形找出拟合残差 $\hat{\varepsilon}_i$ 随 x_i 变化而显示的变化规律，图 10.1 中几种形式都是典型的异方差问题。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

总而言之，异方差就是仅仅违反了假设 3 而遵守其余假设的情况，同方差与异方差的区别可由图 10.2 表示。

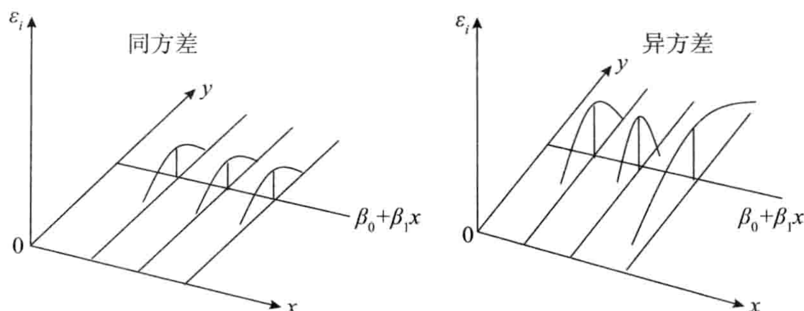


图 10.2

现在，我们所接触到的直观印象已经够多的了，其实异方差的发现还有另外的一些检验方法。这些检验方法能够使我们对付比较复杂的异方差现象。随着学习的深入，我们将知道图示法很多时候是无法解决问题的，比如自回归现象发生时，其方差分布也很有规律性。因此，我们在这里将要介绍的五种有关异方差的检验就变得尤为重要了。

一、RESET 检验（RESET – Test）

这一检验方法是由 F. J. Anscomb 和 J. B. Ramsey 分别于 1961 年、1969 年提出的。它的步骤为：

（一）对线性回归模型进行回归，得到 $\hat{\varepsilon}_i$ 和 $\hat{y}_i$，并由此得到 $\hat{y}_i^2,\hat{y}_i^3,\cdots$，再把 $\hat{\varepsilon}_i$ 分别对 $\hat{y}_i,\hat{y}_i^2,\hat{y}_i^3,\cdots$ 进行回归。

（二）判断每一次回归得到的回归系数是否显著，并由此来建立 $\hat{\varepsilon}_i$ 与 y_i 之间的关系，消去异方差。

RESET 检验的目的，就是要把异方差转化为同方差，从而使得古典模型理论能再一次发挥作用，最小二乘估计量达到最好。因此，RESET 检验力图找到一个与 $\hat{\varepsilon}_i$ 有关的变量 (Z_i)，并找出 $\hat{\varepsilon}_i$ 与 Z_i 之间的相互关系，$\hat{\varepsilon}_i = f(Z_i)$。这个 Z_i 可能是 x_i，也可能是 y_i，或者是 $x_i^2,\hat{y}_i^2,\hat{y}_i^3,\cdots$，各种可能性都存在。为了达到这个目的，它首先要对 $y_i = \alpha + \beta_1 x_{1i} + \cdots + \varepsilon_i$ 进行回归，而不问是否有异方差存在，回归后得到估计残差 $\hat{\varepsilon}_i$ 及 $\hat{y}_i$，再看有无异方差（用 GRAPH 命令）。如果有，接着进行处理。这里 RESET 检验在其步骤中已经向我们暗示，我们只须沿 $\hat{y}_i,\hat{y}_i^2,\hat{y}_i^3,\cdots$ 这一变量序列寻找与 ε_i 有关的变量 Z_i 就足够了。不用再去费力考

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

虑 X_i 等其它变量关系，也就是说，他认为 $\hat{\varepsilon}_i$ 与 $\hat{y}_i, \hat{y}_i^2, \hat{y}_i^3, \cdots$ 之间必然存在一个相关关系。

在得到了 $\hat{y}_i$ 后，再进行 $\hat{\varepsilon}_i$ 对 $\hat{y}_i$ 的回归并考察该回归系数是否显著。如果显著，说明 $\hat{y}_i$ 对 $\hat{\varepsilon}_i$ 影响大，也即 $\hat{y}_i$ 与 $\hat{\varepsilon}_i$ 存在相互影响的关系。如果不显著，再看 $\hat{\varepsilon}_i$ 与 $\hat{y}_i^2$ 的回归系数显著与否，直到找到 $\hat{\varepsilon}_i$ 与 $\hat{y}_i$ 某次方的关系显著为止。

在得到 $\varepsilon_i = \sigma Z_i$ 以后，就可以对一元线性回归模型 $y_i = \alpha + \beta x_i + \varepsilon_i$ 作如下处理：

将模型两边同时除以 Z_i，可得：$\frac{y_i}{Z_i} = (\frac{1}{Z_i})\alpha + \beta(\frac{x_i}{Z_i}) + (\frac{\varepsilon_i}{Z_i})$，那么 $Var(\frac{\varepsilon_i}{Z_i}) = \sigma^2$ 。

这样就可将一个异方差模型化成古典模型来解决了。下面，我们讨论一下如何使用 Eviews 得到 y_i 的拟合值 $\hat{y}_i$ 。我们可以使用 FORECAST 命令来处理，此时计算机屏幕将显示：what is the Series Name? 键入：YH，计算机会自动计算出 y_i 。或用命令 GENR $YH = \hat{\alpha} + \hat{\beta} x_i$ 来得到 $\hat{y}_i$ 。在此之后，再去找出 $\hat{\varepsilon}_i$ 与 $\hat{y}_i$ 的关系。

注意： 实际中造成异方差往往有两个原因：一种情况是模型形式选择上的错误，例如实际生成过程为 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i}^2 + \eta_i$，而回归时误用了 $y_i = \alpha + \beta_1 x_{1i} + \varepsilon_i$，显然 $\varepsilon_i = \beta_2 x_{2i}^2 + \eta_i$ 是一个异方差问题。另一种情况是符合我们前面有关异方差六条假设前提下的真正异方差问题。即，在符合异方差六条假设前提下，没有发生模型形式选择错误的问题。

下面，我们仅就真正的异方差问题进行讨论。

二、White 检验（White – Test）

这种方法由 H. White 于 1980 年提出。它的步骤是：

（一）对方程 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$ 进行回归得到。

（二）$\hat{\varepsilon}_i^2$ 对所有解释变量（即自变量 x_i）及其平方、交叉乘积分别进行回归，并判断回归系数是否显著，并由此建立起 ε_i 与 x_i 之间的相互关系，消去异方差。

这种方法相对来说简单一些，举例来说，假如回归模型有三个自变量：x_{1i}, x_{2i}, x_{3i}，那么 $\hat{\varepsilon}_i^2$ 分别对 x_{1i}, x_{2i}, x_{3i}，$x_{1i}^2, x_{2i}^2, x_{3i}^2$，$x_{1i}x_{2i}, x_{2i}x_{3i}, x_{1i}x_{3i}$ 进行回归，判断哪一个回归系数显著，借此确定 ε_i 与 x_i 的相互关系。

三、Glejster 检验（Glejster – Test）

这种检验由 H. Glejster 于 1969 年提出，其中心思想是 ε_i 仅与自变量 x_i 有联系，是自

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

变量 x_i 的函数，即 $|\hat{\varepsilon}_i| = f(x_i)$ 。Glejster 检验的步骤如下：

（一）对原模型进行回归分析，得出 $\hat{\varepsilon}_i$ 。

（二）$|\hat{\varepsilon}_i|$ 对自变量 x_i 进行回归。

由于该回归模型的实际形式常为未知，Glejster 提出了如下几种形式：

$$|\hat{\varepsilon}_i| = \alpha + \beta x_i$$

$$-|\hat{\varepsilon}_i| = \alpha + \frac{\beta}{x_i}$$

$$\vdots \qquad \vdots$$

（三）对 $H_0: \beta = 0$ 进行检验，判断回归系数显著与否。

在我们上面提到的三种检验中有一个共同的思路，就是认为 ε_i 背后受到一个因素的影响，并存在着一种函数关系，所以总要力图找到一个影响 ε_i 的函数形式。在这三种检验中，实际上背后都隐含了一个假设，$Var(\varepsilon_i) = \sigma^2 f(Z_i)$，其中 Z_i 是未知的。因而不同的检验对 $f(Z_i)$ 有不同的表示方法。

四、Goldfeld & Quandt 检验（Goldfeld & Quandt – Test）

这个检验采用另一种思路，它不涉及有关 $f(Z_i)$ 的具体函数形式而直接对异方差进行检验。这是由 S. M. Goidfeld 和 R. E. Quandt 于 1972 年提出的，使用了我们已经熟知的 F 检验法。

在介绍其检验步骤前先复习一下 F 检验。我们以下面这个例子为蓝本进行复习。

例 对于 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2I} + \beta_3 x_{3i} + \varepsilon_i$，检验假设 $H_0: \begin{cases} \beta_1 = 3 \\ \beta_2 = \beta_3 \end{cases}$ 是否成立。

这时 F 检验步骤为：

1. 对 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$ 回归，得到 $ESS(u)$，而且 $\frac{ESS(u)}{\sigma_u^2} \sim \chi^2_{n-k-1}$；

2. 代入约束条件 $\beta_1 = 3, \beta_2 = \beta_3$，得到 $y_i - 3x_{1i} = \alpha + \beta_2(x_{2i} + x_{3i}) + \varepsilon_i$ 。

令 $y_i^* = y_i - 3x_{1i}, x_i^* = x_{2i} + x_{3i}$，有新方程：$y_i^* = \alpha + \beta_2 x_i^* + \varepsilon_i$，回归后得到 $ESS(R)$，并有 $\frac{ESS(R)}{\sigma_R^2} \sim \chi^2_{n-(k-r)-1}$（此处 $r = 2$，即假设 H_0 中有两个约束）。

3. 计算 F 值：$F_0 = \frac{[ESS(R) - ESS(u)]/r}{ESS(u)/n-k-1} \sim F_{r,n-k-1}$ 。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



4. 把 F_0 与 $F_{r,n-k-1}$ 临界值比较来决定是拒绝还是接受假设 H_0。

这里 F 检验的前提条件是：（1）在新旧两个方程中 ε_i 是相同的；（2）两次回归所使用的数据集是相同的；（3）所要检验的方程具有不同的函数形式。

因此，上面的检验仅对不同函数形式进行检验，而 ε_i 是同样的，则 $\frac{ESS(u)}{\sigma_R^2}$ 与 $\frac{ESS(u)}{\sigma_u^2}$ 相除时可使用：$\sigma_R^2 = \sigma_u^2$。

下面我们要接触的 F 检验是另一种思路了。它的前提条件变了：（1）函数形式相同；（2）两次回归所使用的数据集不同；（3）对残差的方差进行检验：σ_1^2 是否等于 σ_2^2。

这种 F 检验的步骤为：

1. 使用模型 $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$，对数据集进行回归，可以得到数据集 I，的回归残差平方和，记作 ESS_1，并且有：$\frac{ESS(u)}{\sigma_1^2} \sim \chi^2_{n_1-k-1}$，其中有 n_1 为第一数据集中样本个数。

2. 使用同样的方程，对数据集 Ⅱ 进行回归，可得到数据集 Ⅱ 的残差平方和，记作 ESS_2，并且 $\frac{ESS(u)}{\sigma_2^2} \sim \chi^2_{n_2-k-1}$，其中 n_2 为第二数据集中样本个数。

3. 计算 F 值：$F_0 = \frac{(ESS_1/\sigma_1^2)(n_1-k-1)}{(ESS_2/\sigma_2^2)(n_2-k-1)} \sim F_{n_1-k-1,n_2-k-1}$；

这样代入假设 $H_0:\sigma_1^2 = \sigma_2^2$，有：$F_0 = \frac{ESS_1/n_1-k-1}{ESS_2/n_2-k-1} \sim F_{n_1-k-1,n_2-k-1}$。

4. 如果 F_0 大于临界值 F_{n_1-k-1,n_2-k-1}，则拒绝假设 $H_0:\sigma_1^2 = \sigma_2^2$，为异方差，否则接受假设 $H_0:\sigma_1^2 = \sigma_2^2$，是同方差。

我们在了解 F 检验的这一思路后，可以归纳出 Goldfeld & Quandt 检验的步骤来：

（一）把整个数据分成两个数据子集，一组中自变量 x_i 值较大，一组中自变量 x_i 值较小。

（二）运用 F 检验方法对假设 $H_0:\sigma_1^2 = \sigma_2^2$ 进行检验。如果 F_0 > 临界值 F_{n_1-k-1,n_2-k-1}，则拒绝 $H_0:\sigma_1^2 = \sigma_2^2$，说明出现异方差，否则接受假设 $H_0:\sigma_1^2 = \sigma_2^2$，说明没有异方差问题。

除我们上面已讨论过的四个检验外，对异方差问题还有其它的一些检验方法，比如：最大似然法（The Maximum Likelihood Test）、Breusch & Pagan 检验和 Bartlett 检验。限于篇幅，这里就不一一介绍了。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第三节　异方差的后果

在本章第二节中我们对异方差进行了检验，这些检验方法为我们提供了一个思路，就是将异方差转化为同方差以后，再利用古典模型进行回归分析。也就是说，在找到异方差变动的原因后，消除了这一变动原因，使其满足古典假设模型的假设，得到的估计量才是最好的无偏估计量。举例来说：对于一元线性回归模型 $y_i = \alpha + \beta x_i + \varepsilon_i$，如果已知 $Var(\varepsilon_i) = f^2(Z_i)\sigma^2$，我们可以将模型两端同除以 $f(Z_i)$，得到

$$\frac{y_i}{f(Z_i)} = \alpha[\frac{1}{f(Z_i)}] + \beta[\frac{x_i}{f(Z_i)}] + \frac{\varepsilon_i}{f(Z_i)}$$

为什么除以 $f(Z_i)$ 呢？这是为了保证 $Var[\frac{\varepsilon_i}{f(Z_i)}] = \frac{Var(\varepsilon_i)}{f^2(Z_i)} = \frac{f^2(Z_i)\sigma^2}{f^2(Z_i)} = \sigma^2$。这样，实际上等于将模型中所有变量都除以一个权数后，再进行最小二乘回归，这种方法叫加权最小二乘法（Weighted Least square Method），可记作 WLS。从上面的分析可以看出，使用加权最小二乘法能够保证所得到的估计量是无偏的、最好的估计量。

我们现在面临的问题是，如果不使用 WLS，而只用普通最小二乘法，那么所得到的最小二乘估计量会有什么结果呢？下面，我们将详细讨论这个问题。

一、普通最小二乘估计量是无偏的，却不是有效的

假设最简单的模型：$y_i = \beta x_i + \varepsilon_i, \varepsilon_i \sim i.i.d.(0, \sigma_i^2)$，已知 $Var(\hat{\varepsilon}_i) = \sigma_i^2 = \sigma^2 Z_i^2$，也就是说，该模型除了方差外，所有其它古典假设的条件都得到了满足，因此，我们要考察使用普通最小二乘法所得到的普通最小二乘估计量 $\hat{\beta}_{LS}$ 的期望、方差。对于一元线性回归模型 $y_i = \beta x_i + \varepsilon_i$ 而言：

$$\hat{\beta}_{LS} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(\beta x_i + \varepsilon_i)}{\sum x_i^2} = \frac{\sum x_i^2\beta + \sum x_i\varepsilon_i}{\sum x_i^2} = \beta + \frac{\sum x_i\varepsilon_i}{\sum x_i^2}$$

$$E(\hat{\beta}_{LS}) = E(\beta + \frac{\sum x_i\varepsilon_i}{\sum x_i^2})$$

$$= \beta + \frac{\sum x_i E(\varepsilon_i)}{E(\sum x_i^2)} \text{（古典假设保证了 } E(\varepsilon_i) = 0\text{）}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$= \beta$$

因此，$E(\hat{\beta}_{LS}) = \beta$。

再看方差，$Var(\hat{\beta}_{LS}) = Var(\beta + \frac{\sum x_i\varepsilon_i}{\sum x_i^2}) = Var(\frac{\sum x_i\varepsilon_i}{\sum x_i^2}) = \frac{\sum x_i^2 Var(\varepsilon_i)}{(\sum x_i^2)^2} = \frac{\sum x_i^2\sigma_i^2}{(\sum x_i^2)^2}$。

这样使用普通最小二乘法得到的普通最小二乘估计量的期望和方差为：

$$E(\hat{\beta}_{LS}) = \beta$$

$$Var(\hat{\beta}_{LS}) = \frac{\sum x_i^2\sigma_i^2}{(\sum x_i^2)^2}$$

如果用加权最小二乘法得到的估计量的方差和期望又是怎样呢？

假设一元线性回归模型：$y_i = \beta x_i + \varepsilon_i, \varepsilon_i \sim i.i.d.(0,\sigma_i^2)$，已知 $Var(\varepsilon_i) = \sigma_i^2 = Z_i^2\sigma_i^2$。

在方程两边同除以 Z_i 后，得 $\frac{y_i}{Z_i} = \beta(\frac{x_i}{Z_i}) + (\frac{\varepsilon_i}{Z_i})$。

这样，我们就得到了一个新的一元线性回归模型，我们利用这个新的一元回归模型得到的估计量就是加权最小二乘估计量 $E(\hat{\beta}_{WLS})$，则

$$\hat{\beta}_{WLS} = \frac{\sum(\frac{x_i}{Z_i})(\frac{y_i}{Z_i})}{\sum(\frac{x_i}{Z_i})^2} = \frac{\sum[\beta(\frac{x_i}{Z_i})^2 + (\frac{\varepsilon_i x_i}{Z_i^2})]}{\sum(\frac{x_i}{Z_i})^2} = \beta + \frac{\sum(\frac{x_i}{Z_i})(\frac{\varepsilon_i}{Z_i})}{\sum(\frac{x_i}{Z_i})^2}$$

$$E(\hat{\beta}_{WLS}) = \beta + \frac{\sum(\frac{x_i}{Z_i})E(\frac{\varepsilon_i}{Z_i})}{\sum(\frac{x_i}{Z_i})^2}\text{（由于 }E(\frac{\varepsilon_i}{Z_i}) = \frac{E(\varepsilon_i)}{Z_i} = 0\text{）}$$

$$= \beta$$

$$Var(\hat{\beta}_{WLS}) = \frac{\sigma^2}{\sum(\frac{x_i}{Z_i})^2}$$

我们是怎么得到 $Var(\hat{\beta}_{WLS})$ 呢？新的一元线性回归模型：$\frac{y_i}{Z_i} = \beta(\frac{x_i}{Z_i}) + (\frac{\varepsilon_i}{Z_i})$ 告诉我们，原模型的异方差已被转化为了同方差。这个新模型满足古典假设的一切条件。因此，可以按最小二乘法对 $y_i^* = \beta x_i^* + \varepsilon_i$ 进行计算，则：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum x_i^*} = \frac{\sigma^2}{\sum \left(\frac{x_i}{Z_i}\right)^2}$$

由此可见，估计量 $\hat{\beta}_{WLS}$ 和 $\hat{\beta}_{LS}$ 都是无偏的，但二者方差却不同。下面，我们就比较一下二者的方差：

$$\frac{Var(\hat{\beta}_{WLS})}{Var(\hat{\beta}_{LS})} = \frac{\dfrac{\sigma^2}{\sum \left(\frac{x_i}{Z_i}\right)^2}}{\dfrac{\sum x_i^2\sigma_i^2}{(\sum x_i^2)^2}} = \frac{\dfrac{\sigma^2}{\sum \left(\frac{x_i}{Z_i}\right)^2}}{\dfrac{\sum x_i^2(\sigma^2 \cdot Z_i^2)}{(\sum x_i^2)^2}} = \frac{\dfrac{\sigma^2}{\sum \left(\frac{x_i}{Z_i}\right)^2}}{\dfrac{\sigma^2 \sum x_i^2 Z_i^2}{(\sum x_i^2)^2}} = \frac{1/\sum \left(\frac{x_i}{Z_i}\right)^2}{\dfrac{\sum x_i^2 Z_i^2}{(\sum x_i^2)^2}} = \frac{(\sum x_i^2)^2}{\sum \left(\frac{x_i}{Z_i}\right)^2 \sum x_i^2 Z_i^2}$$

由于分子中 $\frac{x_i}{Z_i}$ 与 $x_i Z_i$ 的乘积等于 x_i^2，令 $\frac{x_i}{Z_i} = a_i$，$Z_i x_i = b_i$，初等数学知识告诉我们 $\frac{(\sum ab)^2}{\sum a^2 \sum b^2} \leq 1$，显然

$$\frac{(\sum x_i^2)^2}{\sum \left(\frac{x_i}{Z_i}\right)^2 \sum x_i^2 Z_i^2} \leq 1$$

这样，我们得到了如下结论，$Var(\hat{\beta}_{WLS}) \leq Var(\hat{\beta}_{LS})$。这说明普通最小二乘估计量 $\hat{\beta}_{LS}$ 是无效的。

二、$\hat{\beta}_{LS}$ 的方差估计量是有偏的

什么是 $\hat{\beta}_{LS}$ 的方差估计量呢？在古典模型中，我们知道：$\hat{\beta}_{LS} \sim i.i.d\left(\beta, \frac{\sigma^2}{\sum x_i^2}\right)$。

那么，$\frac{\hat{\beta}_{LS} - \beta}{\sqrt{\frac{\sigma^2}{\sum x_i^2}}} \sim i.i.d(0,1)$。

在实际计算中，由于 σ^2 往往是未知的，因此，我们经常使用估计量 $\frac{S^2}{\sum x_i^2}$ 代替 β 的真实方差 $\frac{\sigma^2}{\sum x_i^2}$，这样 $\frac{S^2}{\sum x_i^2}$ 就是 $\hat{\beta}_{LS}$ 真实方差 $\frac{\sigma^2}{\sum x_i^2}$ 的估计量。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

下面，我们要讨论的问题为 $\frac{S^2}{\sum x_i^2}$ 是有偏的。

就 $y_i = \beta x_i + \varepsilon_i$ 来说，$\hat{\beta}_{LS}$ 的方差估计量为 $\frac{S^2}{\sum x_i^2} = \frac{\frac{ESS}{n-k-1}}{\sum x_i^2}(n-k-1=n-1)$，

$E\left(\frac{S^2}{\sum x_i^2}\right) = E\left(\frac{\frac{ESS}{n-1}}{\sum x_i^2}\right) = \frac{E(ESS)}{(n-1)\sum x_i^2}$。

由于 $E(ESS) = E[\sum (y_i - \hat{\beta}_{LS}x_i)^2] = \sum \sigma_i^2 - \frac{\sum x_i^2\sigma_i^2}{x_i^2}$，那么

$$E\left(\frac{S^2}{\sum x_i^2}\right) = \frac{E(ESS)}{(n-1)\sum x_i^2} = \frac{\sum \sigma_i^2 - \frac{\sum x_i^2\sigma_i^2}{x_i^2}}{(n-1)\sum x_i^2} = \frac{\sum x_i^2 \sum \sigma_i^2 - \sum x_i^2\sigma_i^2}{(n-1)\left(\sum x_i^2\right)^2}。$$

总之，由于 $Var(\hat{\beta}) - E\left[\frac{ESS}{(n-1)\sum x_i^2}\right] = \frac{n\left(\frac{1}{n}\sum x_i^2 \sum \sigma_i^2 - \sum x_i^2\sigma_i^2\right)}{(n-1)\left(\sum x_i^2\right)^2}$ 很难等于零，这说明 $\frac{S^2}{\sum x_i^2}$ 是有偏的。在这种情况下，协方差 $\mathrm{cov}(x_i,\varepsilon_i)$ 的不确定性会给我们进行 t 检验带来很大麻烦。在进行 t 检验：$t_0 = \frac{\hat{\beta}_{LS} - \beta}{\sqrt{\frac{S^2}{\sum x_i^2}}}$ 时，如果 σ_i 与 x_i 正相关，会导致估计量 $\frac{S^2}{\sum x_i^2}$ 小于 $\hat{\beta}_{LS}$ 的真实方差，即 $E\left(\frac{S^2}{\sum x_i^2}\right) < Var(\hat{\beta}_{LS})$。这样，在对假设 $H_0: \beta = 0$ 进行检验时，会使得我们可能因为 t 检验值的显著而接受该估计值。如果 σ_i 与 x_i 负相关，则导致估计量 $\frac{S^2}{\sum x_i^2}$ 大于 $\hat{\beta}_{LS}$ 的真实方差，即 $E\left(\frac{S^2}{\sum x_i^2}\right) > Var(\hat{\beta}_{LS})$，从而使我们对假设 $H_0: \beta = 0$ 进行检验时发生偏差，很有可能因为 t 检验值的不显著而拒绝估计量。这就告诉我们，如果不管 β_i 的方差是同是异，而一概简单地使用普通最小二乘法，忽视了 $Var(\hat{\beta}_{LS})$ 估计量 $\frac{S^2}{\sum x_i^2}$ 有偏这一点，将使我们承担以上两种错误的后果。也就是说，我们有可能

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

低估或高估最小二乘估计量 $\hat{\beta}_{LS}$ 的真正方差，这将影响我们对假设 $H_0: \beta = 0$ 进行检验的准确性。

总的来说，解决异方差问题的方法，关键在于我们对异方差产生原因的解释，但当我们无法找到这一产生原因时，我们至少可以尽力去订正 $\hat{\beta}_{LS}$ 方差估计量。既然我们已经知道 $E(\frac{S^2}{\sum x_i^2})$ 本身是有偏的，可不可以干脆使用公式 $Var(\hat{\beta}_{LS}) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2}$ 来算方差呢？White 和 C. R. Rao 分别于 1980 年、1970 年建议，在使用时，由于 σ^2 未知 S^2 有偏，那么就使用回归拟合残差 $\hat{\varepsilon}_i$ 代替 σ_i 计算 $Var(\hat{\beta}_{LS})$。

第四节　异方差的解决方法

解决异方差有两类方法：一类是找出异方差变动的原因后再予以解决；另一类是一般解决办法。下面，我们分别进行讨论。

一、基于有关 σ_i^2 假设上的解决方法

这种解决方法，实际上就是找出异方差变动原因后再对异方差问题进行解决。这里，我们一般可以使用加权最小二乘法和最大似然法（Maximum Likelihood Method）这两种方法来解决。下面对有关加权最小二乘法的内容进行讨论，并列举几种典型或有特点的情况，对其进行分析研究。

第一种情况：对于一元线性回归模型：$y_i = \alpha + \beta x_i + \varepsilon_i$，已知 $Var(\varepsilon_i) = \sigma^2 Z_i^2$。

使用 WLS 解决方法：$\frac{y_i}{Z_i} = \alpha(\frac{1}{Z_i}) + \beta(\frac{x_i}{Z_i}) + (\frac{\varepsilon_i}{Z_i})$ 对新模型进行回归分析，就可以得到无偏有效估计量 $\hat{\alpha}_{WLS}$，$\hat{\beta}_{WLS}$。

第二种情况：对于一元线性回归模型：$y_i = \beta x_i + \varepsilon_i$，同样假设已知 $Var(\varepsilon_i) = \sigma^2 x_i$。

使用 WLS 解决方法，建立新模型：$\frac{y_i}{\sqrt{x_i}} = \beta(\sqrt{x_i}) + \frac{\varepsilon_i}{\sqrt{x_i}}$。

我们对新模型进行回归，能不能解决异方差问题呢？当然可以。但是我们之所以使用 $Var(\varepsilon_i) = \sigma^2 x_i$，是因为此时的 $\hat{\beta}_{WLS}$ 的计算公式将带给我们极大的乐趣：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\hat{\beta}_{WLS} = \frac{\sum\left(\frac{y_i}{\sqrt{x_i}}\right)\sqrt{x_i}}{\sum\left(\sqrt{x_i}\right)^2} = \frac{\sum y_i}{\sum x_i} = \frac{\bar{y}}{\bar{x}}$$

第三种情况：对一元线性回归模型：$y_i = \alpha + \beta x_i + \varepsilon_i$，同样假设已知 $Var(\varepsilon_i) = \sigma^2 x_i$。

在使用 WLS 法时，会得到新模型：$\frac{y_i}{x_i} = \alpha(\frac{1}{x_i}) + \beta + (\frac{\varepsilon_i}{x_i})$。

这里原来模型的系数 β 成了新模型中的常数项，而原来的常数项 α 却成了新模型中的系数。

第四种情况：对一元线性回归模型：$y_i = \alpha + \beta x_i + \varepsilon_i$，已知 $Var(\varepsilon_i) = \sigma^2[E(y_i)]^2 = \sigma^2(\alpha + \beta x_i)^2$。

这是一个十分特别的情况：$\alpha + \beta x_i$ 是未知量，如何导出 $Var(\varepsilon_i)$ 呢？下面我们将介绍两步加权最小二乘法（Two Stage WLS Method），这种方法，由 Paris 和 Houthhakle 于 1955 年提出。它的基本思路是：既然不知道 $\alpha + \beta x_i$，就用 $\hat{\alpha}$、$\hat{\beta}$ 代替。

其具体步骤如下：

（一）对 $y_i = \alpha + \beta x_i + \varepsilon_i$ 直接用普通最小二乘法进行回归，得到估计量 $\hat{\alpha}$、$\hat{\beta}$、y 的拟合值 $\hat{y}$，其中 $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$。

（二）采用加权最小二乘法（WLS 法）。将原模型两边都除以 $\hat{y}_i$，可得：

$$\frac{y_i}{\hat{y}_i} = \alpha(\frac{1}{\hat{y}_i}) + \beta(\frac{x_i}{\hat{y}_i}) + (\frac{\varepsilon_i}{\hat{y}_i})$$

对这个新方程进行回归，可以得到 α、β 的估计值 $\hat{\alpha}_{WLS}$，$\hat{\beta}_{WLS}$。

总之，加权最小二乘法（WLS）的目的就是在异方差下得到有效、无偏估计量。而此时普通最小二乘法的估计量由于方差过大而失效。WLS 法为达到这一目的，首先找出异方差产生原因，然后把异方差转化为了同方差。这样，古典模型的六个假设都能够得以满足，因此，此时再使用普通最小二乘法就能够得到一个最好、无偏的估计量。

二、一般解法

一般解法其实也分两类：一类是设法将变量除以某些“标准”，得到消除了异方差后的变量方程；另一类是把数据先转化为对数形式，再进行回归。在计量经济实践中，计量经济学家很偏爱使用对数解决问题，往往一开始就是把数据转化为对数形式，再对对数方程进行回归。这主要因为对数形式可以减少异方差和自相关的程度。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



当我们把变量除以某些“标准”，得到消除了异方差的新方程时，就是说明我们使用了加权最小二乘法（WLS）。一般来说，这里的“标准”往往比加权最小二乘法（WLS）要宽得多。加权最小二乘法要求的是方程两边所有的变量都必须除以同一个“标准”，而这里所讨论的一般解法中的“标准”，并不一定像加权最小二乘法那样严格，很有可能只是方程中某一变量或某几个变量除以某一“标准”，同样可以解决异方差的问题。然而，这里讨论的一般解法中的两种方法也会带来一些令人迷惑的问题。对第一种解法来说，会产生新经济含义和假象相关问题；对第二种解法来说，会产生线性形式与对数形式选择比较的问题。下面我们将分别讨论之。

（一）新经济含义和假象相关的产生和解决

假设一个某地政府要为市民修建一个公共设施，例如健身休闲广场，建设费用为该地区的财权支出的一部分。

TC 表示总成本（Total Cost），M 表示广场平米数。用 x 表示均摊到每平米建设的投入量（Input），可以得到方程

$$TC = \alpha \cdot M + \beta \cdot x + \varepsilon$$

这里的 ε 包括了一切建设广场时可能出现的其他成本。当然，如果把固定成本（Fixed Cost）考虑进去，建立的 TC 方程可能是这样的：

$$TC = \alpha \cdot M + \beta \cdot x + FC + \varepsilon$$

其中，FC 表示固定成本，是一个常数。

在解这个模型时，你可能会发现这个模型含有异方差问题，而你将决定采用加权最小二乘法来解决，以两边同除以 M 来抵消异方差，于是 $TC = \alpha \cdot M + \beta \cdot x + \varepsilon$ 变成了

$$\frac{TC}{M} = \alpha + \beta \frac{x}{M} + \frac{\varepsilon}{M}$$

同样，$TC = \alpha \cdot M + \beta \cdot x + FC + \varepsilon$ 变成 $\frac{TC}{M} = \alpha + \beta \frac{x}{M} + FC \frac{1}{M} + \frac{\varepsilon}{M}$，后一个方程无非是想说，在变形中，原来是常数的 FC 变成了变量的系数，而作为变量系数的 α 却成了常数。如果这种做法可以消除异方差，那么，$\frac{TC}{M}$ 就有了新的经济含义。这里 $\frac{TC}{M}$ 表示了一个比率变量（Ratio Variable）——单位面积成本。问题就在于此，我们在使用包括加权最小二乘法在内的“变量除以标准”方法，本意是要消除异方差，但在这一过程中，却产生了新的经济变量。怎么办呢？所有处理异方差的过程，其目的是为了得到普通最小二乘法无法得到的更为有效的估计量。一旦得到这些估计量，我们将使用变形前的原方程来进行预

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



测和推断，也就是说，尽管在消除异方差时变量有时可能产生新的经济含义，我们一概不管。只要能求得到该新方程的无偏估计量，该估计量在未变形的原方程中也是最有效的，可以用它并以原方程为依据，进行预测和推断。

假象相关是怎么产生的呢？也许 x_i 与 y_i 相关度可能不高，但是在除以一个标准 Z_i 后，$(\frac{x_i}{Z_i})$ 和 $(\frac{y_i}{Z_i})$ 可能相关度会马上提高，这就是所谓的**假象相关**（Spurious Correlation）。这同样也就说明，由于 y_i 和 x_i 的原始模型得出的 R^2，不能同 $\frac{y_i}{Z_i}$ 与 $\frac{x_i}{Z_i}$ 的模型得出的 R^2 相比，此时仅通过 R^2，无法说明哪个方程拟合得更好。

假象相关是 1955 年由 Kuh 和 Meyer 提出。他们认为，尽管 y_i 和 x_i 不相关，但如果将两个变量同除以另一个变量 Z_i，在 $\frac{x_i}{Z_i}$ 与 $\frac{y_i}{Z_i}$ 之间就可能有很强的相关程度。这是不是意味着这种估计方法就不能使用了呢？举例来说，$y_i = \alpha Z_i + \beta x_i + \gamma M_i + \varepsilon_i$，$\varepsilon_i \sim i.i.d(0,\sigma_i^2)$，已知异方差生成规律 $Var(\varepsilon_i) = \sigma^2 Z_i^2$，仍使用加权最小二乘法，可得：

$$(\frac{y_i}{Z_i}) = \alpha + \beta(\frac{x_i}{Z_i}) + \gamma(\frac{M_i}{Z_i}) + (\frac{\varepsilon_i}{Z_i})$$

上式中 $\frac{y_i}{Z_i}$、$\frac{x_i}{Z_i}$、$\frac{M_i}{Z_i}$ 三个变量的相关程度很可能很高，并可能会导致多重共线性。这时必须记住，我们除以某个“标准”，只是为了得到最有效的估计量 α、β、γ，即我们关心的是 α、β 和 γ 无偏估计量，而对那些纠缠不清的假象相关问题没有多少兴致，只要在一定的能够容忍的范围内，即相关度不等于 1（实际上是最多只能接近于 1，而无法等于 1），我们进行回归，仍能得出无偏的参数估计量 $\hat{\alpha}$、$\hat{\beta}$、$\hat{\gamma}$。这就是说，只要我们不去比较 $y_i = \alpha Z_i + \beta x_i + \gamma M_i + \varepsilon_i$ 和 $(\frac{y_i}{Z_i}) = \alpha + \beta(\frac{x_i}{Z_i}) + \gamma(\frac{M_i}{Z_i}) + (\frac{\varepsilon_i}{Z_i})$ 的 R^2 值，那么，假象相关存在与否，对我们求无偏估计量无关；$\text{cov}(x_i,y_i)$ 是大于、等于，还是小于 $\text{cov}(\frac{x_i}{Z_i},\frac{y_i}{Z_i})$ 都无足轻重。

（二）线性形式与对数形式的选择

经济计量学家们热衷于对对数形式的线性方程进行回归分析，是因为对数形式能够减少异方差、自相关现象，以致于在数据输入后，往往先把数据直接转化为对数形式，再进行回归。然而有时用了对数函数后，却仍然有可能产生异方差问题，在这种情况下，将迫

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

使我们重新回过来思考异方差产生的原因。因为，此时异方差很可能已无法用对数形式消除了。BM 检验等检验方法将帮助我们在此时作出正确选择：对于线性形式还是对数形式回归，哪一种方法能把异方差处理得更好。

考虑一个生产函数，我们很可能使用线性形式：

$$Y = \alpha + \beta_1 L + \beta_2 K$$

其中，L 、K 分别为劳动与资本，Y 为产出。

也可能使用对数形式：

$$(\log Y) = \alpha + \beta_1(\log L) + \beta_2(\log K)$$

但应当注意，这两个函数具有不同的性质：

$Y = \alpha + \beta_1 L + \beta_2 K$ 中的 L 和 K 是完全可以替代的。少用 L，就可以通过增加 K 来解决。这种完全替代（Perfect Substitution）关系，我们可用图 10.3 表示。

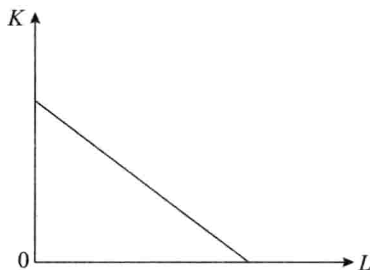


图 10.3

有完全替代，自然就有完全不可替代（Perfect Unsubstitution），著名的里昂惕夫生产函数就是一种，即

$$Y = \mathrm{Min}(\frac{L}{a},\frac{K}{b})$$

这里 $\mathrm{Min}(\frac{L}{a},\frac{K}{b})$ 表示在 $\frac{K}{b}$ 、$\frac{L}{a}$ 中只取最小的那个。现实中的形象说明，两个人看一台机床，如果有 100 台机床（$K=100$），有 200 人（$L=200$），则 $Y = \mathrm{Min}(\frac{200}{2},\frac{100}{1}) = 100$。如果 $K=200$，$L=300$，则 $Y = \mathrm{Min}(\frac{300}{2},\frac{200}{1}) = 150$。这种完全不可替代关系，我们可用图 10.4 表示：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



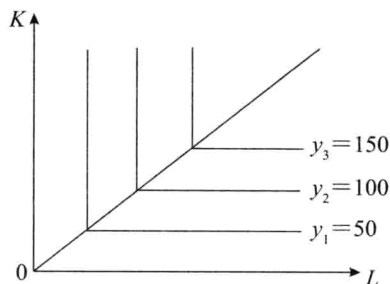


图 10.4

介于完全替代和完全不可替代之间的自然就是不完全替代（Unperfect Substitution）了，如图 10.5 所示：

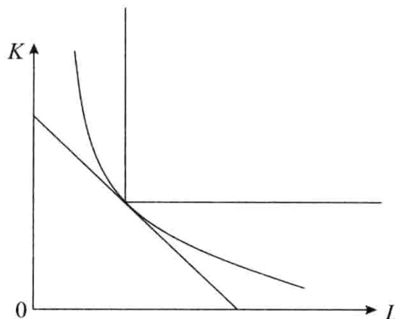


图 10.5

柯布–道格拉斯函数就是一个最好的不完全替代的例子，其方程式为：

$$(\log Y) = \alpha + \beta_1(\log L) + \beta_2(\log K)$$

在描述经济系统运行中，有时会用到柯布–道格拉斯生产函数，有时会用里昂惕夫生产函数，有时还可能用线性生产函数。选择是多种多样的，问题是选择哪一种好。

再如需求方程：$Q = \alpha + \beta_1 P + \beta_2 Y$。其中，$Q$ 表示需求量，P 为价格，Y 为收入。

采用这种线性需求函数的形式，就意味着需求弹性是变化的。若采用对数形式的需求方程：

$$(\log Q) = \alpha + \beta_1(\log P) + \beta_2(\log Y)$$

则意味着需求弹性是不变的。它比线性模型更为直观，因为它的 $\hat{\beta}_1$ 就是价格弹性，$\hat{\beta}_2$ 就是收入弹性。撇开经济涵义不谈，我们现在讨论是采用对数形式还是采用线性形式；线性形式所得到的 R^2 能否与对数形式下得到的 R^2 相比。当然，这两个 R^2 是不能相比的，通

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

过上面两个不同形式方程对比可以看出，*TSS* 变了，前者为 Y，后者是 $\log Y$ 。

下面介绍两种对对数形式和线性形式模型异方差的检验方法：

1. BM 检验

BM 检验由 Bera 和 McAleer 于 1982 年提出。

假设要检验的对数形式和线性形式模型已经给出：

$H_0:\log Y_t = \alpha + \beta_1 x_t + u_t, u_t \sim i.i.d\ N(0,\sigma_0^2)$

$H_1:Y_t = \alpha + \beta_1 x_t + u_t, u_t \sim i.i.d\ N(0,\sigma_1^2)$

这里是要检验使用哪一个形式更好一些。

BM 检验的步骤：

（1）无论要采用哪一个，都需对上面两个方程进行回归，并得到了拟合值（$\log\hat{y}_t$）和 $\hat{y}_t$ 。

（2）分别对如下方程进行回归：

$a.\ Exp(\log\hat{y}_t) = \alpha + \beta_1 x_t + V_{1t}$，得到 $\hat{V}_{1t}$ ；

$b.\ \log\hat{y}_t = \alpha + \beta_1 x_t + V_{2t}$，得到 $\hat{V}_{2t}$ 。

（3）分别对以下方程回归：对 H_0，H_1 的选择取决于 θ_0，θ_1 。

$\log\hat{Y}_t = \alpha + \beta_1 x_t + \theta_0 \hat{V}_{1t} + \varepsilon_t$

$\hat{Y}_t = \alpha + \beta_1 x_t + \theta_1 \hat{V}_{2t} + \varepsilon_t$

（4）进行 t 检验。在对（3）中的两个方程进行回归时，计算机会给出 θ_0，θ_1 系数的 t 检验值。如果 θ_0 的 t 检验值不显著，说明 $\hat{V}_{1t}$ 对 $\log Y_t$ 没什么影响，此时接受假设 H_0，即选择对数形式；如果 θ_0 的 t 检验值显著，则看 θ_1 的检验值，如果 θ_1 的 t 检验值不显著，则选择线性形式。这便是 BM 检验的全部过程。

用 BM 检验也有问题：如果 θ_0，θ_1 的 t 检验值都很显著，或者 θ_0，θ_1 的 t 检验值都不显著，则 BM 检验将无法判断。

2. PE 检验

PE 检验由 Mackinnon，White 和 Davidson 于 1983 年提出。

PE 检验的步骤是：

（1）同 BM 检验第一步。

（2）对下面两个方程进行回归分析：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



$$\log(y_t) = \alpha + \beta_1 x_t + \theta_0[\hat{y}_t - Exp(\log\hat{y}_t)] + \varepsilon_t$$

$$y_t = \alpha + \beta_1 x_t + \theta_1[(\log\hat{y}_t) - \log\hat{y}_t] + \varepsilon_t$$

检验 $\theta_0 = 0$，$\theta_1 = 1$ 的可能性。

（3）同 BM 检验第四步。

实际上 PE 检验的思路同 BM 检验是一样的，它只不过用了不同的方程形式而使步骤更为简单了。

除此之外，还有需要使用最大似然估计的 Box – Cox 检验等多种方法，帮助我们正确地选择方程形式。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第十一章 自相关

第一节 自相关概述

什么是自相关？在第十章我们已经描述了自相关的假设前提是：

假设 1　随机项纵向变动

假设 2　$E(\varepsilon_i) = 0$

假设 3　$Var(\varepsilon_i) = \sigma^2$（注意是同方差）

假设 4　$\text{cov}(\varepsilon_i, \varepsilon_j) \neq 0$（此处是唯一违反古典假设前提的地方）

假设 5　$\text{cov}(x_i, x_j) = 0$；$\text{cov}(\varepsilon_i, x_j) = 0$

假设 6　线性生成过程：$y_i = \alpha + \beta x_i + \varepsilon_i$

在 ε_i，ε_j 相关时有两种情况：第一种是**截面数据**（Cross Section Data）下的 ε_i，ε_j 相关。在回归方程中，这一影响将会反映到 ε_i 中去，就引起了自相关，但是截面数据中这种 ε_i 和 ε_j 的相关关系，不叫自相关，而叫**空间相关**（Spatial Correlation）。它的解决办法是运用虚拟变量。第二种是时间序列数据下的 ε_i，ε_j 相关。它将引起自相关（Autocorrelation），又称**序列相关**（Serial Correlation）。

序列相关是指残差项 ε_t 与 $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots$ 相关，其中 ε_t 与 ε_{t-k} 的相关叫作 k **阶自相关**。那么：

$$\rho_{\varepsilon_t, \varepsilon_{t-1}} = \rho_1 = \frac{\text{cov}(\varepsilon_t, \varepsilon_{t-1})}{\sqrt{Var(\varepsilon_t)} \cdot \sqrt{Var(\varepsilon_{t-1})}} \text{ 称一阶自相关；}$$

$$\rho_{\varepsilon_t, \varepsilon_{t-2}} = \rho_2 = \frac{\text{cov}(\varepsilon_t, \varepsilon_{t-2})}{\sqrt{Var(\varepsilon_t)} \cdot \sqrt{Var(\varepsilon_{t-2})}} \text{ 称二阶自相关；}$$

$$\rho_{\varepsilon_t, \varepsilon_{t-k}} = \rho_k = \frac{\text{cov}(\varepsilon_t, \varepsilon_{t-k})}{\sqrt{Var(\varepsilon_t)} \cdot \sqrt{Var(\varepsilon_{t-k})}} \text{ 称 } k \text{ 阶自相关。}$$

本章将讨论两个问题：1. 如何对序列相关进行检验；2. 关于序列相关下的估计方法。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第二节 Durbin - Watson 检验

最简单、最常用的模型是 ε_t 与 ε_{t-1} 一阶自相关系数为 ρ 的自相关模型。

用方程表示为：

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

$$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + v_t,\ v_t \sim i.i.d.(o, \sigma^2)$$

由于 ρ 是母体中的，无法得到，因此，在这一模型中，我们在考虑用 $\hat{\rho}$ 代替 ρ 的基础上对 ρ 进行假设检验，其中：

$$\hat{\rho} = \frac{\operatorname{cov}(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1})}{\sqrt{Var(\hat{\varepsilon}_t)} \cdot \sqrt{Var(\hat{\varepsilon}_{t-1})}}$$

这里最常用的统计量为 Durbin - Watson 统计量，用 d 来表示：

$$d = \frac{\sum_{i=1}^{n-1}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{i=1}^{n}\hat{\varepsilon}_t^2} = \frac{\sum_{i=1}^{n-1}\hat{\varepsilon}_t^2 - 2\sum\hat{\varepsilon}_t\hat{\varepsilon}_{t-1} + \sum\hat{\varepsilon}_{t-1}^2}{\sum_{i=1}^{n}\hat{\varepsilon}_t^2}$$

对于大样本来说，$\hat{\varepsilon}_t$ 是可以得到的，且 $\hat{\varepsilon}_t^2$ 与 $\hat{\varepsilon}_{t-1}^2$ 相差很小，不妨认为是相等的。由于 $\rho \doteq \frac{\sum\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{\sum\hat{\varepsilon}_t^2}$，因此，$d \doteq 2 - 2\rho \doteq 2(1-\rho)$，如果 $\rho = 0$，则 $d \doteq 2$ 。

因为 ρ 为相关系数，且 $-1 \le \rho \le 1$，那么有 $0 \le d \le 4$ 。

计算机软件在我们回归后会自动告诉我们 DW 检验值。如果 $d = 2$，或者 d 极接近 2，则不存在自相关的问题。由于 d 是个随机变量，因此，它有自己的分布。因此，Durbin 和 Watson 针对 d 的显著水平规定了上限（用 du 表示）、下限（用 dL 表示），如图 11.1 所示。

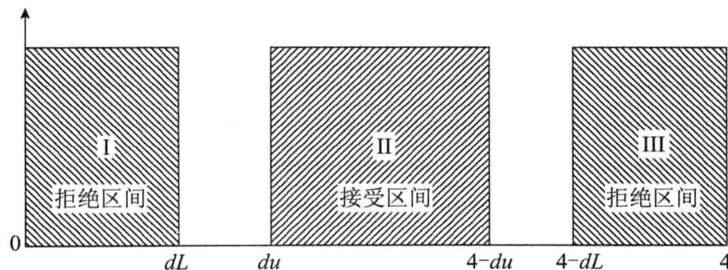


图 11.1

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



如果 $d \in [0,2]$ 之间，ρ 为大于零的正相关；如果 $d \in [2,4]$，ρ 为小于零的负相关。关于 DW 检验值有以下三个区域：

1. 如果 DW 检验值 $d \in (dL, du)$ 或 $d \in [4-du, 4-dL]$ 时，无法判断是否自相关。
2. 如果 DW 检验值 $d \in (du, 4-du)$，即区域Ⅱ，不存在自相关问题。
3. $d \in [0, dL]$ 或 $d \in [4-dL, 4]$ 时，即在区域Ⅰ、Ⅲ区域分别为一阶正、负自相关。

上面我们说 $d \doteq 2(1-\rho)$，这是对大样本而言。实际上当 $\rho = 0$ 时，d 的期望为：

$$E(d) = 2 + \frac{2(k-1)}{n-k}$$

自然有 $\mathrm{plim}E(d) = 2$，但当 n 较小时，$E(d)$ 会偏离 2。

Durbin - Watson 检验也有如下缺限：

1. 它仅对一阶自相关有效。
2. 当 Durbin - Watson 检验值 $d \in (dL, du)$ 或 $d \in (4-du, 4-dL)$ 时，无法确定是否存在自相关。
4. 该检验无法解释含有滞后因变量的方程。如：$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 x_t + \varepsilon_t$ 。

第三节　自相关的处理方法

高斯-马尔科夫定理所指出的最小二乘估计量最好这一结论，是在古典模型的六个假定条件都得到满足的情况下成立的。如果其它假定满足而有关自相关的假定被破坏，普通最小二乘估计量就不再是有效估计量了。其估计方差将变大，参数置信区间将变宽，t 检验效力下降，普通最小二乘估计量无效。这和出现异方差时所面临的问题很相似。

同样，遵循类似思路，我们设法找到了测量自相关的标准，这就是本章第二节所述的 DW 检验值 d 。d 统计量有一个基本前提，就是随机扰动项 ε_t 之间的关系服从一阶自相关，即 $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ 。这是个十分重要的假设前提。本节所讲的所有内容都是建立在其基础上的。统计量 d 值告诉我们，在 du 到 $4-du$ 区间内不存在自相关问题。换言之，如果我们计算得到的 DW 检验值很接近 2，我们使用普通最小二乘法将得到一个有效无偏估计量。但是如果 d 值落在了 $(0, du)$ 和 $(4-du, 4)$ 里，又该怎么办呢?

回想一下，我们是如何处理异方差的。解决异方差时，我们使用了加权最小二乘法（WLS），它的手段之一就是将原有方程加以变形，成为不含异方差的新方程，进而得到无偏有效估计量。这里也将如法炮制：我们将设法构造出一个新方程来，这个新方程是在原

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

有方程基础上变形得到的，并且消除了自相关，成为完美地符合古典假设条件的模型。

对于模型

$$y_t = \alpha + \beta x_t + \varepsilon_t,$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t, v_t \sim i.i.d.(o, \sigma_v^2)$$

这里 v_t 符合古典假设，v_t 的方差为 $\sigma_v{}^2$。这样做是为了区别 ε_t 的方差 σ^2 而命名的，它的脚码仅表示是 v_t 的方差，而不是异方差的变动符。

如果，$y_t = \alpha + \beta x_t + \varepsilon_t$ (11.1)

在 t 时刻成立，则它在 $t-1$ 时刻也自然成立，所以有：

$$y_{t-1} = \alpha + \beta x_{t-1} + \varepsilon_{t-1} \tag{11.2}$$

上式两边同乘以 ρ，有：

$$\rho y_{t-1} = \alpha\rho + \beta\rho x_{t-1} + \rho\varepsilon_{t-1} \tag{11.3}$$

将（11.1）式 -（11.3）式，则有：

$$y_{t-1} - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + \varepsilon_t - \rho\varepsilon_{t-1}$$

而 $\varepsilon_t - \rho\varepsilon_{t-1} = v_t$，则以上方程变为：

$$y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + v_t$$

这是一个不带自相关的符合古典假设的方程，对其进行回归，得到的 $\hat{\alpha}$、$\hat{\beta}$ 是 α、β 的有效无偏估计量。

至此一切变得很清楚了：如果 DW 检验拒绝了 $H_0: \rho = 0$ 的假设，也就是说 DW 检验值 d 不接近于 2，证明有自相关的存在。此时由于 ε_t 自相关，普通最小二乘估计量将无效。要解决这一问题，就需要把不产生自相关的随机项 v_t 引进方程中，取代 ε_t 从而使新方程消除了自相关。

方程 $y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + v_t$ 称广义的差分模型或准差分模型。当 $\rho = 1$ 时，将是标准的一阶差分模型。然而，真正的 ρ 是很难知道的，在实践中我们往往用 $\hat{\rho}$ 来代替。$\hat{\rho}$ 又是从何得来的呢？它的求解如同 DW 检验中 $\hat{\rho}$ 的求解方法一样，从 $y_t = \alpha + \beta x_t + \varepsilon_t$ 中得到 $\hat{\varepsilon}_t$，那么，在大样本的情况下：$\hat{\rho} = \dfrac{\sum \hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_t^2}$。

$\hat{\rho}$ 作为 ρ 的估计量，会受到样本残差 $\hat{\varepsilon}_t$ 的限制，所以一般人们使用另外的方法，即：如果 d 与 0 或 4 的距离很小时，将直接使用一阶差分方程，如用 $y_t - y_{t-1} = \beta(x_t - x_{t-1}) + v_t$ 来回归，实际上是令 $\rho = 1$。经验告诉我们：只要 DW 的检验值不显著，就使用一阶差分方程。例如，模型 $y_t = \alpha + \beta x_t + \varepsilon_t$ 存在自相关，则可对方程 $y_t - y_{t-1} = \beta(x_t - x_{t-1}) + v_t$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

进行回归分析。这是典型的一阶差分方程，可令 $y_t^* = y_t - y_{t-1}, x_t^* = x_t - x_{t-1}$，该方程转化为：$y_t^* = \beta x_t^* + v_t$。

这时候可以直接使用最小二乘法进行回归，但是值得注意的是，我们将难以从该方程中得到 α 的估计值，当然，并非所有差分方程都没常数项。请看下面的例子：

$$y_t = \alpha + \beta t + \gamma x_t + \varepsilon_t \tag{11.4}$$

$$y_{t-1} = \alpha + \beta(t-1) + \gamma x_{t-1} + \varepsilon_{t-1} \tag{11.5}$$

将（11.4）式 -（11.5）式得

$$y_t - y_{t-1} = \beta + \gamma(x_t - x_{t-1}) + (\varepsilon_t - \varepsilon_{t-1})$$

常数项出现了。

与差分方程相区别，我们通常称原始的方程为水平方程，如 $y_t = \alpha + \beta t + \gamma x_t + \varepsilon_t$。

在计算机 EViews 中进行差分计算，需进行如下操作：如要产生 $y_t^* = y_t - y_{t-1}$ 这样的数列，可以使用 GENR 命令：GENR YTS = Y - Y(-1)。

必须当心，在 GENR 命令使用前，要用 SMPL 命令修改一下样本范围。例如，y_t 序列表示 1979—1983 年的数据值，y_{t-1} 实际是向下错出一位后横向相减得到的。例如图 11.2 所示：

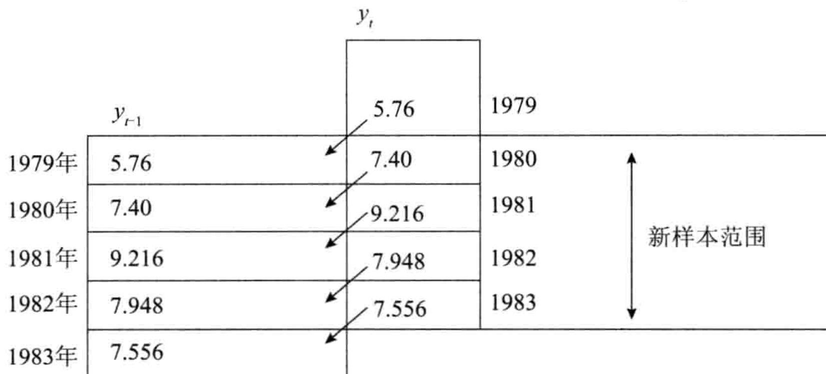


图 11.2

这样 y_t 与 y_{t-1} 横向相减，就可以表达出 $y_t - y_{t-1}$ 的值：

$$y_1^* = 7.40 - 5.76$$

$$y_2^* = 9.216 - 7.40$$

$$y_3^* = 7.948 - 9.216$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



$$y_4^* = 7.556 - 7.948$$

我们有了差分方程，当然要比较一下差分方程与水平方程的优劣。但是，我们不能直接用两类方程的 R^2 来比较说明，因为被解释变量在两类方程中是不一样的。换言之，TSS 是不同的。为了衡量这两种方程的优劣，我们可以对 ESS 进行粗略调整后再进行比较。在本节中我们都是以差分方程的指标为基准，调整水平方程的相应指标后，再进行相互比较的。

对于 $y_t = \alpha + \beta x_t + \varepsilon_t$，$Var(\varepsilon_t) = \sigma^2$ 而言，回归后可得水平方程的 ESS，记作 ESS_{level}，其自由度为 $n - k - 1$ 。

对于 $y_t - y_{t-1} = \beta(x_t - x_{t-1}) + (\varepsilon_t - \varepsilon_{t-1})$ 回归后可得差分方程的 ESS，记作 ESS_{first} 。自由度为 $n - k$，$Var(\varepsilon_t - \varepsilon_{t-1}) = Var(\varepsilon_t) + Var(\varepsilon_{t-1}) - 2\text{cov}(\varepsilon_t, \varepsilon_{t-1})$ 。

如果在 t 时刻，有：

$$y_t = \alpha + \beta x_t + \varepsilon_t, Var(\varepsilon_t) = \sigma^2$$

在 $t-1$ 时刻，有：

$$y_{t-1} = \alpha + \beta x_{t-1} + \varepsilon_{t-1}, Var(\varepsilon_{t-1}) = \sigma^2$$

而 $\text{cov}(\varepsilon_t, \varepsilon_{t-1})$ 在本章第二节中已经说过为 $\rho\sigma^2$ ，因此，$Var(\varepsilon_t - \varepsilon_{t-1}) = \sigma^2 + \sigma^2 - 2\rho\sigma^2 = 2(1-\rho)\sigma^2$ 。

可见，如果要对 ESS_{level} 进行调整，首先要调整自由度，其次还要调整方差。我们用 ESS_{level}^D 表示可以同 ESS_{first} 相比较的 ESS_{level}，那么：

$$ESS_{level}^D = ESS_{level} \times \underbrace{\left(\frac{n-k-1}{n-k}\right)}_{\text{调整自由度}} \times \underbrace{2(1-\rho)}_{\text{调整方差}}$$

这样 ESS_{level}^D 就可以同 ESS_{first} 相比较了。

又因为 $2(1-\rho) \doteq d$，这里 d 是水平方程的 DW 检验值，不妨记作 d_{level}，则：

$$ESS_{level}^D = ESS_{level} \times \left(\frac{n-k-1}{n-k}\right) \times d_{level}$$

当然，有时候使用 R^2 进行比较更为方便。但两类方程的 R^2 无法直接相比，我们不妨效仿 ESS 的调整方法，对 R^2 也如法炮制，调整之后再比较，仍以差分方程的 R_{first}^2 为基准。我们定义：R_L^2 表示可与一阶差分方程的 R_{first}^2 相比的水平方程的 R^2；R_{first}^2 表示一阶差分方程的 R^2 ；ESS_{level} 表示来自水平方程的 ESS；ESS_{first} 表示来自一阶差分方程的 ESS。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

由 $\frac{1-R_D^2}{1-R_{first}^2}=\frac{ESS_{level}\times(\frac{n-k-1}{n-k})\times d_{level}}{ESS_{first}}=\frac{ESS_{level}}{ESS_{first}}\times(\frac{n-k-1}{n-k})\times d_{level}$

可得 $R_D^2=1-(1-R_{first}^2)\cdot\frac{ESS_{level}}{ESS_{first}}(\frac{n-k-1}{n-k})\cdot d_{level}$

下面将举一个美国生产函数的例子以加深我们的印象：

美国生产函数：$\log x_t=-3.938+1.451\log L_1+0.384\log k_1$。已知 $R_{level}^2=0.9946$，DW 检验 $d=0.858$，$ESS_{level}=0.0434$，样本个数为 39。

对于 $\Delta\log x_t=0.987\times\Delta\log L_1+0.502\times\Delta\log k_1$

这里：$\Delta\log x_t=\log x_t-\log x_{t-1}$，$\Delta\log L_1=\log L_{1t}-\log L_{1t-1}$，$\Delta\log k_1=\log k_{1t}-\log k_{1t-1}$。

又得到 $R_{first}^2=0.8405$，DW 检验 $d=1.177$，$ESS_{first}=0.0278$，仔细分析一下上面的两个方程及其结论，我们会发现：如果对 R^2 不经调整而直接比较，R_{level}^2 显然高于 R_{first}^2，这说明水平方程比差分方程更令人满意。然而，再看 DW 检验值，DW_{first} 又比 DW_{level} 好一些。但这是一种表面现象，我们需要调整 R_{level}^2，然后再进行比较。

$$R_D^2=1-(1-0.8405)\cdot\frac{0.0434}{0.0278}\cdot\frac{36}{37}\cdot 0.858=0.7921$$

$R_D^2<R_{first}^2$ 说明差分方程更好一些。

此外，哈韦（Harvey）曾经给出了一个定义不同的 R_D^2：

$$R_D^2=1-\frac{ESS_{level}}{ESS_{first}}\cdot(1-R_{first}^2)$$

很显然，他没有考虑自由度和方差的调整。这样可能出现的问题是：$|R_D^2|\leq 0$。因此 R_D^2 很有可能是负值。

对于一般的时间序列数据来说，我们总能发现这样的规律：从表面上看，$R_{level}^2>R_{first}^2$，但如果 DW 检验值很小，就应该使用差分方程而不是使用水平方程，这叫做 R^2 综合症（R^2 Syndrome）。

再次重申一下，我们目前为止所接触到的各种指标，诸如 DW 检验值、ESS_{level}、R_D^2、一阶差分方程，都是针对 $\varepsilon_t=\rho\varepsilon_{t-1}+v_t$ 一阶自相关 AR（1）而言的。如果 $\varepsilon_t=\rho_1\varepsilon_{t-1}+\rho_2\varepsilon_{t-2}+v_t$ 或 $\varepsilon_t=\rho\varepsilon_{t-2}$，这是二阶自相关 AR（2）的问题，此时就不能再使用上述指标了。

时间序列中季度性数据往往是 $\varepsilon_t=\rho\varepsilon_{t-4}$ 的 4 阶自相关，今年某一季度与去年同季度的经济行为相联系；月度数据可能有 $\varepsilon_t=\rho\varepsilon_{t-12}$ 是 12 阶自相关。因此，只有年度数据是一阶自相关 $\varepsilon_t=\rho\varepsilon_{t-1}$ 可以接受上述指标。同时，一阶以上的自相关，其 DW 检验值要重新定义。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第四节　自相关误差下的回归过程

对于 $y_t = \alpha + \beta x_t + \varepsilon_t$，$\varepsilon_t = \rho\varepsilon_{t-1} + v_t, v_t \sim i.i.d.(o,\sigma^2)$ 来说，是典型的一阶自相关 AR（1）。

$$\begin{aligned}\sigma^2 &= Var(\varepsilon_t)\\ &= Var(\rho\varepsilon_{t-1} + v_t)\\ &= \rho^2 Var(\varepsilon_{t-1}) + Var(v_t) + 2\rho\text{cov}(\varepsilon_{t-1}, v_t)\end{aligned}$$

由于 $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ 符合古典假设，因此，$\text{cov}(\varepsilon_{t-1}, v_t) = 0$；同样，由于时间性，$Var(\varepsilon_{t-1}) = \sigma^2$，那么，$\rho^2 Var(\varepsilon_{t-1}) + Var(v_t) + 2\rho\text{cov}(\varepsilon_{t-1}, v_t) = \rho^2\sigma^2 + \sigma_v^2$。

则：$\sigma^2 = \dfrac{\sigma_v^2}{1-\rho^2}$。

同时，对于 $y_t = \alpha + \beta x_t + \varepsilon_t$

$$\begin{aligned}E(\varepsilon_t\varepsilon_{t-s}) &= E[(\rho\varepsilon_{t-1} + v_t)\varepsilon_{t-s}]\\ &= E(\rho\varepsilon_{t-1}\varepsilon_{t-s} + v_t\varepsilon_{t-s})\\ &= \rho E(\varepsilon_{t-1}\varepsilon_{t-s}) + E(v_t\varepsilon_{t-s})\\ &= \rho\cdot\rho_{s-1}\end{aligned}$$

如果，$\rho_0 = 1$，那么 $\rho_1 = \rho$，$\rho_2 = \rho^2$，$\rho_3 = \rho^3, \cdots$，$\rho_s = \rho^s$，只要 $\rho < 1$，$\rho_s = \rho^s$ 将是收敛级数，如图 11.3 所示：

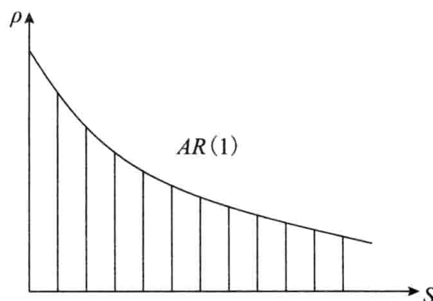


图 11.3

图 11.3 表示的是典型的一阶自相关 AR（1）。在 EViews 软件中，可用 IDENT RESID

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



命令看到它。命令输入后，计算机会问你：Number of Lags? 通常令 S = 10，键入 10 并回车后，即可得到屏幕输出的结果。

一、广义最小二乘法（General Least Square Method，简称 GLS）

对于 $y_t = \alpha + \beta x_t + \varepsilon_t, \varepsilon_t = \rho\varepsilon_{t-1} + v_t, (t = 1, 2, 3, \cdots, T)$ 来说

$$\rho y_{t-1} = \alpha\rho + \beta\rho x_{t-1} + \rho\varepsilon_{t-1}$$

两个方程相减，可得：$y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + v_t$。

令 $y_t^* = y_t - \rho y_{t-1}, \alpha^* = \alpha(1-\rho), x_t^* = x_t - \rho x_{t-1}$，则：$y_t^* = \alpha^* + \beta x^* + v_t$，其中，$t = 2, 3, \cdots, T$。

在这种广义差分法下，差分以后丢了一个观测值，因为第一个观测值是没有先行值的。为了避免这个观测值的损失，应该使用以下方法定义一对 y 和 x 的第一个观测值，并把它们补充到样本中：

$$x_1^* = x_1\sqrt{1-\rho^2}$$

$$y_1^* = y_1\sqrt{1-\rho^2}$$

在这种情况下，再进行的最小二乘法回归，是广义最小二乘法的回归。实际上，这个 ρ 往往很难得到，我们总是设法找到 $\hat{\rho}$ 来代替 ρ。寻找 $\hat{\rho}$ 有两种方法：循环查找法（Iterative Procedure）和灰色寻找法（Grid - Search Procedure）。

二、循环查找法

循环查找法有两个过程，科克伦 - 欧卡特查找过程和杜宾过程。下面分别介绍：

（一）科克伦 - 欧卡特循环查找过程（Cochran - Orcutt Procedure）

1. 对 $y_t = \alpha + \beta x_t + \varepsilon_t$ 进行回归，得到 $\hat{\varepsilon}_i$ 和 ESS，$ESS_{旧} = ESS$。

2. 计算 $\hat{\rho} = \dfrac{\sum \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_t^2}$。

3. 对 $y_t - \hat{\rho} y_{t-1} = \alpha(1-\hat{\rho}) + \beta(x_t - \hat{\rho} x_{t-1}) + v_t$ 进行回归，得到 $\hat{\alpha}$、$\hat{\beta}$，并代入步骤 1 的方程中求出 ESS，令 $ESS_{新} = ESS$。

4. 如果 $\left|\dfrac{ESS_{新} - ESS_{旧}}{ESS_{旧}}\right| < 0.05$，$\rho$ 即为所要得到的估计值；否则：

（1）令 $\hat{\varepsilon}_t = y_t - (\hat{\alpha} + \hat{\beta} x_t)$；

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

（2）$ESS_{新}$的值赋予 $ESS_{旧}$。

5. 重复步骤 2 及以下程序，直到 $\left|\frac{ESS_{新}-ESS_{旧}}{ESS_{旧}}\right|<0.05$ 得到满足为止。

这时 $\hat{\rho}$ 是使 ESS 最小的 ρ 的估计值。

必须证明的是，这种方法是以 ESS 收敛为前提的。计算机上的 Eviews 软件包就是采用这种方法得到 $\hat{\rho}$ 。在肯定是一阶自相关的条件下，可以键入命令：

LS　Y　C　X　AR（1）

计算机屏幕将首先显示 $\hat{\rho}$ 的寻找过程，然后自动进行 GLS 法回归。在结果中 AR（1）一项的系数就是所要寻找的 $\hat{\rho}$ 的值。

（二）杜宾循环查找过程（Durbin Procedure）

1. 首先对 $y_t=\alpha(1-\rho)+\rho y_{t-1}+\beta x_t-\beta\rho x_{t-1}+v_t$ 进行回归，得到 $\hat{\rho}$ 和 ESS，令 $ESS_{旧}=ESS$。

2. 对 $y_t-\hat{\rho}y_{t-1}=\alpha(1-\hat{\rho})+\beta(x_t-\hat{\rho}x_{t-1})+v_t$ 进行回归，得到 $\hat{\alpha}$、$\hat{\beta}$，并代入步骤 1 的方程中求出 ESS，令 $ESS_{旧}=ESS$。

3. 如果 $\left|\frac{ESS_{新}-ESS_{旧}}{ESS_{旧}}\right|<0.05$，$\rho$ 即为所要得到的估计值；否则：

（1）对 $y_t=\hat{\alpha}(1-\rho)+\rho y_{t-1}+\rho x_t-\hat{\beta}x_{t-1}+v_t$ 进行回归；

（2）令 $ESS_{旧}=ESS_{新}$。

4. 重复 2、3，直到 $\left|\frac{ESS_{新}-ESS_{旧}}{ESS_{旧}}\right|<0.05$ 条件得到满足为止。

由于加入了新变量 y_{t-1}，此法并非常用。

三、灰色寻找法

灰色寻找法由希尔德斯（Hildreth）和卢（Lu）于 1960 年提出。具体步骤如下：

1. 在 $-1\leq\rho\leq1$ 范围内每间隔 0.1 选一个 ρ。

2. 对 $y_t-\rho y_{t-1}=\alpha(1-\rho)+\beta(x_t-\rho x_{t-1})+v_t$ 进行回归，得出 ESS。

3. 选出使 ESS 最小的 ρ 来。

这个方法的思路是，由于不知道 ρ 的确定值，只知道 $-1\leq\rho\leq1$，就分别取 $\rho=-1$，$\rho=-0.9$，$\rho=-0.8$，$\rho=-0.7$，…，$\rho=0.9$，$\rho=1$，每取一次 ρ 值，做一次回归，并得到相应的 ESS，从中进行比较，找出使 ESS 最小的 ρ 来。当然可能有如下情况：比如当 ρ =

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

0.8 和 $\rho=0.7$ 时 *ESS* 是一样的，都是最小的，那么就把这个区间再分成十个小区间，令 $\rho=0.71$，$\rho=0.72$，…，$\rho=0.79$，重复步骤 2，直到区别出使 *ESS* 最小的 ρ 为止。

第五节　一阶自相关对最小二乘估计量的影响

就像讨论异方差对最小二乘估计量的影响一样，我们要找出这种影响，并且留心估计量期望和方差的变化，判断普通最小二乘估计量是否是有效的、无偏的、最好的估计量。

典型的一阶自相关 AR（1）模型最简单的形式如下：

$$y_t = \beta x_t + \varepsilon_t, \varepsilon_t = \rho\varepsilon_{t-1} + v_t$$

对之进行最小二乘估计，有：$\hat{\beta} = \dfrac{\sum x_t y_t}{\sum x_t^2}$。

把 $y_t = \beta x_t + \varepsilon_t$ 代入上式，可以推出：

$$\begin{aligned}\hat{\beta} &= \frac{\sum x_t(\beta x_t + \varepsilon_t)}{\sum x_t^2} \\ &= \frac{\sum x_t^2\beta}{\sum x_t^2} + \frac{\sum x_t\varepsilon_t}{\sum x_t^2} \\ &= \beta + \frac{\sum x_t\varepsilon_t}{\sum x_t^2}\end{aligned}$$

那么，$E(\hat{\beta}) = \beta + \dfrac{\sum x_t E(\varepsilon_t)}{\sum x_t^2}$。

我们假设 $y_t = \beta x_t + \varepsilon_t$ 中 ε_t 是自相关，但并没否认 $E(\varepsilon_t)=0$，因此：$E(\hat{\beta})=\beta$。这说明在一阶自相关下，普通最小二乘估计量是无偏的。

再看方差

$$\begin{aligned}Var(\hat{\beta}) &= Var(\beta + \frac{\sum x_t\varepsilon_t}{\sum x_t^2}) \\ &= Var(\frac{\sum x_t\varepsilon_t}{\sum x_t^2})\text{（由于 }\beta\text{ 为常数，}Var(\beta)=0\text{）}\end{aligned}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



$$= \frac{Var(\sum x_t\varepsilon_t)}{(\sum x_t^2)^2}$$

$$= \frac{1}{(\sum x_t^2)^2}[\sum x_t^2 Var(\varepsilon_t) + 2\sum x_t x_{t-1}\text{cov}(\varepsilon_t,\varepsilon_{t-1})$$

$$+ 2\sum x_t x_{t-2}\text{cov}(\varepsilon_t,\varepsilon_{t-2}) + \cdots]（由于 \varepsilon_t 为自相关）$$

前面已经证明过：

$$\frac{\text{cov}(\varepsilon_t,\varepsilon_{t-1})}{Var(\varepsilon_t)} = \rho_1 = \rho^1$$

$$\frac{\text{cov}(\varepsilon_t,\varepsilon_{t-2})}{Var(\varepsilon_t)} = \rho_2 = \rho^2$$

$$\vdots$$

那么：

$$\text{cov}(\varepsilon_t,\varepsilon_{t-j}) = \sigma^2\rho^j$$

$$Var(\hat{\beta}) = \frac{1}{(\sum x_t^2)^2}(\sigma^2\sum x_t^2 + 2\sigma^2\rho\sum x_t x_{t-1} + 2\sigma^2\rho^2\sum x_t x_{t-2} + 2\sigma^2\rho^3\sum x_t x_{t-3} + \cdots)$$

$$= \frac{\sigma^2}{\sum x_t^2}(1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + 2\rho^3\frac{\sum x_t x_{t-3}}{\sum x_t^2} + \cdots)$$

此时我们假设 x_t 一阶自相关，即：$x_t = \Phi x_{t-1} + \eta_t$，且 $Var(x_t) = \sigma^2$，$\text{cov}(x_t,x_{t-1}) = \Phi$，则：

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum x_t^2}(1 + 2\rho\Phi + 2\rho^2\Phi^2 + 2\rho^3\Phi^3 + \cdots)$$

$$= \frac{\sigma^2}{\sum x_t^2}(1 + \frac{2\rho\Phi}{1-\rho\Phi})（由于 \rho 为小于 1 的数）$$

$$= \frac{\sigma^2}{\sum x_t^2}(\frac{1+\rho\Phi}{1-\rho\Phi})$$

这时我们所得到的 $\hat{\beta}$ 的方差，是根据真实情况，考虑了自相关并调整以后得到的。如果我们贸然对 $y_t = \beta x_t + \varepsilon_t$ 进行最小二乘回归，有 $Var(\hat{\beta}_{LS}) = \frac{\sigma^2}{\sum x_t^2}$。那么：

$$Var(\hat{\beta}) = (\frac{1+\rho\Phi}{1-\rho\Phi})Var(\hat{\beta}_{LS})$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

计算机在进行回归时，给我们显示出来的就是 $Var(\hat{\beta}_{LS})$ 。我们可以看出，$Var(\hat{\beta}_{LS}) = (\frac{1-\rho\Phi}{1+\rho\Phi})Var(\hat{\beta})$ 。这证明未调整值 $Var(\hat{\beta}_{LS})$ 会远远小于真实的方差。这样的后果就是，参数 $\hat{\beta}$ 的 t 检验值和 F 检验值会很大，因而显得很显著，从而使我们盲目接受。举例来说，假如有 $\rho = \Phi = 0.8$，则 $Var(\hat{\beta}_{LS}) = \frac{1-0.64}{1+0.64} \cdot Var(\hat{\beta}) = 0.2Var(\hat{\beta})$ 。

这样计算出的方差缩小了很多，因此，t 检验值和 F 检验值会扩大很多倍而失真。

综上所述，在自相关存在的情况下，贸然地使用最小二乘法，有如下后果：

1. 无偏但无效。

2. 因为 $\hat{\beta}$ 的方差低估，t 检验值和 F 检验值及 R^2 会被夸大。在异方差的情况下，比最小二乘法更有效的估计可以构造出来。在自相关的情况下，普通最小二乘法估计是一致和渐近正态的，因此，由于相关造成的最大的实际问题是由于有偏方差估计而产生出来导致无效的结论，甚至大样本中也是如此。

$Var(\hat{\beta}_{LS})$ 和 $Var(\hat{\beta})$ 什么时候才相等呢？当 $\Phi = 0$ 或 $\rho = 0$ 时，相等的情况就会出现。t 检验、F 检验和 R^2 值才是可信的。

在计量经济学中，要解决自相关，可以使用广义最小二乘法（GLS）、最大似然法（ML）或其它一些方法。但是，在运用这些方法解决自相关时，需要注意以下几点：

1. 如果 ρ 已知，使用 GLS 法就能得到更好的 $\hat{\beta}$ 。

如果 ρ 已知，从 $y_t = \alpha + \beta x_t + \varepsilon_t$，$\varepsilon_t = \rho\varepsilon_{t-1} + v_t$，$\rho y_{t-1} = \rho\alpha + \rho\beta x_{t-1} + \rho\varepsilon_{t-1}$ 中，得到 $y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + v_t$ 符合古典假设，α 、β 的最小二乘估计量最好。

如果 ρ 未知，有时我们就未必能从估计值 $\hat{\rho}$ 中受益。饶（Rao）和吉利彻（Gilicher）在 1969 年提出，在一般的 DW 检验中，如果由 d 得到的 ρ，其绝对值小于等于 0.3 时，在样本数小于等于 20 的情况下，使用寻找法来得到 $\hat{\rho}$ 是毫无益处的。而在其它情况下使用寻找法则可以得到很好的 $\hat{\rho}$ 来协助我们解决问题，见表 11－1。

表 11－1

$\hat{\rho}$ 值 \ 样本数	≤20	>20
$\lvert\hat{\rho}\rvert > 0.3$	使用寻找法	使用寻找法
$\lvert\hat{\rho}\rvert \le 0.3$		使用寻找法

2. 如果 ε_t 的结构更为复杂，如 AR（2）（二阶自相关）、MA（1）（移动平均），若贸

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



然地将 ε_t 作为一阶自相关来处理，只会引起更坏的结果。

3. 对于季度数据和月度数据来说，即使我们可以从 DW 检验中得到不是一阶自相关的结论，也不意味着不存在自相关问题。比如，对于季度数据来说，也许就存在着 $\varepsilon_t = \rho\varepsilon_{t-4} + v_t$。

第六节　对含滞后因变量的序列相关模型的检验

典型的含滞后自变量的序列相关模型为：

$$y_t = \alpha y_{t-1} + \beta x_t + \varepsilon_t, \varepsilon_t = \rho\varepsilon_{t-1} + v_t \tag{11.6}$$

这个模型中的滞后变量为 y_{t-1}。我们知道，在 $y_t = \alpha + \beta x_t + \varepsilon_t, \varepsilon_t = \rho\varepsilon_{t-1} + v_t$ 中，没有 y_t 与 y_{t-1} 的自相关问题，相关的只是 ε_t。而将上式滞后一期，可得：

$$y_{t-1} = \alpha y_{t-2} + \beta x_{t-1} + \varepsilon_{t-1} \tag{11.7}$$

这证明 y_{t-1} 也是随机变量，ε_{t-1} 影响 y_{t-1}，进而影响 y_t。因此，y_{t-1} 与 y_t 之间并非独立的。这也就意味着 $y_t = \alpha y_{t-1} + \beta x_t + \varepsilon_t$ 中有双重自相关。当样本数很大时，说明 $\hat{\alpha}$、$\hat{\beta}$ 是有偏的：$\text{plim}(\hat{\alpha}) = \alpha + A$，$\text{plim}(\hat{\rho}) = \rho - A$，其中 $A = \dfrac{\rho\sigma_*^2\sigma^2}{(1-\alpha\rho)D^2}$，$D = Var(y_{t-1})Var(x) - \text{cov}^2(y_{t-1}, x)$。

这时我们的结论是：带滞后变量的序列相关模型，使得 DW 检验已不再适用。因为 DW 检验只为古典假设中有一条不符的情况服务，而当方程多处违反古典假设时，DW 检验将是无效的。于是杜宾（Durbin）于 1970 年提出 h 检验专门对付这一问题。

h 检验是检验假设 $H_0: \rho = 0$。其步骤如下：

1. 直接回归 $y_t = \alpha y_{t-1} + \beta x_t + \varepsilon_t$，得到 $\hat{\rho}$、$\hat{\alpha}$ 等来计算 $h = \hat{\rho}\sqrt{\dfrac{n}{1 - nVar(\hat{\alpha})}} \sim N(0,1)$。当置信度为 α 时，$h > N_{\frac{\alpha}{2}}$，就拒绝假设 $H_0: \rho = 0$；如果 $h < N_{\frac{\alpha}{2}}$，接受假设 $H_0: \rho = 0$。以此来判断自相关是否存在。

2. 如果 $1 - nVar(\hat{\alpha}) < 0$，那么，应首先对 $y_t = \alpha y_{t-1} + \beta x_t + \varepsilon_t$ 进行回归，得到 $\hat{\varepsilon}_t$；其次，对 $\hat{\varepsilon}_t = \alpha^*\hat{\varepsilon}_{t-1} + \beta^* y_{t-1} + \gamma^* x_t$ 进行回归；再次，检验 α^* 是否等于 0，看 t 检验值即可。并通过此检验来检验 ρ 是否等于 0。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第七节 DW 检验的更为深入的结论

DW 检验是自相关测度上一大重要工具。以上的分析已经向我们揭示了 DW 检验并非像我们开始接触时那样简单了。如果进行更为复杂的研究，对于这个令人感到高深莫测的检验方法就会引出更为深入的结论。

结论一：对于经济数据来说，使用 DW 检验时，用上限边界 du 作为显著点更好。

例如：$\alpha = 5\%$，$n = 25$，$k = 4$，$dL = 1.04$，$du = 1.77$，如果 $d = 1.5$，甚至 $d = 1.70$ 或 1.76，对于经济数据来说，也应最好作自相关处理。这也就是建议我们在 DW 检验中 $[0,4]$ 区间限度内，最好应把 $[du, 4 - du]$ 范围外的一切区域都看作自相关并作相应处理。

结论二：DW 检验的是置信度一般为 0.05 或 0.01。

可以说，如果我们把研究自相关作为最终目的，高的置信度是较好的。但是如果我们的最终目的是预测，处理自相关仅仅是一种必不可少的过程，那么就不一定要求过高的置信度。一般置信度为 30%、40% 就很不错了。

结论三：DW 检验的显著性可能是由多种因素造成的。

如：一阶自相关，也可能是二阶、三阶自相关，变量丢失、动态误定等，所有这些情况我们将在第八节中详细讨论。

第八节 DW 显著时的策略

DW 检验只适宜一阶自相关 AR（1），而对 AR（2），AR（3），…，MA（1），MA（2），…，ARMA（1，1），…就无能为力了。这里 MA 指移动平均，ARMA 则是移动平均叠加自相关。

在 DW 检验值显著时，也就是它远离 2 时（这一点与 t 检验、F 检验定义的显著性正好相反），我们讨论如下产生原因：

（1）非一阶自相关因素；

（2）丢失变量；

（3）动态误定。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



一、残差结构非一阶自相关

我们处理自相关的方法，无非是找到自相关的阶数，以及决定自相关的形式：是自相关（AR），移动平均（MA），自相关移动平均混合（ARMA），还是分布滞后自相关（PDL）。这两个步骤哪个应先考虑，高德利（Hendry）和特雷维迪（Trivedi）提出他们的取舍见解：自相关的阶数要比其形式更重要，对经济变量而言，只须沿着 AR 寻找阶数就足够了。

二、自相关由丢失变量引起

若数据真正生成过程是 $y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t$，我们用 $y_t = \beta_0 + \beta_1 x_t + v_t$ 来预测，那么 $v_t = \beta_2 x_t^2 + \varepsilon_t$，$x_t$ 的自相关导致 v_t 变成了自相关。

若数据真正生成过程是 $y_t = \beta_1 x_t + \beta_2 z_t + \varepsilon_t$，我们用 $y_t = \beta_1 x_t + \varepsilon_t$ 来估计。尽管真正模型符合古典假设，但由于丢失自相关变量 z_t，故而估计模型有严重自相关。

那么，在这种情况下，由于变量丢失，DW 检验值 $\hat{d}$ 会有什么后果呢？

假设 $z_t = bx_t + \eta_t, \mathrm{cov}(\eta_t, \eta_{t-1}) = \rho^* \sigma_*^2$，这也就是说，先找到丢失的 z_t 与 x_t 的关系，并将其代入到真正的方程中：

$$y_t = \beta_1 x_t + \beta_2(bx_t + \eta_t) + \varepsilon_t = (\beta_1 + \beta_2 b)x_t + (\beta_2 \eta_t + \varepsilon_t)$$

从 $y_t = \beta_1 x_t + v_t$ 中得到的最小二乘估计量 $\hat{\beta}_{LS}$ 是有偏的，期望为：$E(\hat{\beta}_{LS}) = \beta_1 + \beta_2 b$。

由于 $\rho = \dfrac{\mathrm{cov}(v_t, v_{t-1})}{Var(v_t)}$，

$$\mathrm{cov}(v_t, v_{t-1}) = \mathrm{cov}(\beta_2 \eta_t + \varepsilon_t, \beta_2 \eta_{t-1} + \varepsilon_{t-1}) = \beta_2^2 \rho^{*2} \sigma_*^2,$$

$$Var(v_t) = Var(\beta_2 \eta_t + \varepsilon_t) = \beta_2^2 \sigma_*^2 + \sigma^2,$$

则：$\hat{\rho} = \dfrac{\beta_2^2 \rho^{*2} \sigma_*^2}{\beta_2^2 \sigma_*^2 + \sigma^2} = \dfrac{\rho^{*2}}{1 + \dfrac{\sigma^2}{\beta_2^2 \sigma_*^2}}$。

由此可见，$d = 2(1 - \rho)$ 的决定因素除了 z_t、x_t 的残差的自相关系数 ρ^* 外，还有 σ^2、z_t 与 y_t 的回归系数 β_2，以及 z_t 与 x_t 回归残差的方差 σ_*^2 的影响。

我们可以用 RESET 检验来检验丢失变量。如果 DW 检验和 RESET 检验都很显著，我们就应加入一些变量，而不使用一阶自相关。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

三、动态误定引起的序列相关

对于数据产生过程 $y_t = \beta x_t + \varepsilon_t, \varepsilon_t = \rho\varepsilon_{t-1} + v_t$ 来说，使用 GLS 后，会有 $y_t - \rho y_{t-1} = \beta(x_t - \rho x_{t-1}) + v_t$。这一方程相当于对 $y_t = \rho y_{t-1} + \beta x_t - \beta\rho x_{t-1} + v_t$ 进行回归。而模型 $y_t = \beta_1 y_{t-1} + \beta_2 x_t + \beta_3 x_{t-1} + v_t$ 是个稳定动态模型，与 $y_t = \rho y_{t-1} + \beta x_t - \beta\rho x_{t-1} + v_t$ 不是同一个模型。

$y_t = \rho y_{t-1} + \beta x_t - \beta\rho x_{t-1} + v_t$ 与 $y_t = \beta_1 y_{t-1} + \beta_2 x_t + \beta_3 x_{t-1} + v_t$ 只有在 $\beta_1\beta_2 + \beta_3 = 0$ 时才会相等。因此，在二者等价的情况下，对 $\rho = 0$ 进行检验就是对 $\beta_1 = 0$(和 $\beta_3 = 0$）进行检验。但是，萨根（Sargan）在 1964 年提出：首先应对限定条件 $H_0: \beta_1\beta_2 + \beta_3 = 0$ 进行检验，如果 H_0 不被拒绝，再检验 $\rho = 0$ 与否。如果 H_0 被拒绝，DW 检验值显著就是由动态误定引起的，即数据的产生过程是：$y_t = \beta_1 y_{t-1} + \beta_2 x_t + \beta_3 x_{t-1} + v_t$。但是 F 检验是无法验证 $\beta_1\beta_2 + \beta_3 = 0$ 的。针对这一问题需要使用的是 Walt 检验，限于篇幅，这里就不再详述了。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第十二章
联立方程模型

我们从最原始的古典假设条件中，已逐步地讨论了自相关、异方差以及多重共线性等问题。对于模型 $y = \alpha + \beta x + \varepsilon$，要满足古典假设，就要符合六个条件，如果仅有假设 $\text{cov}(x,\varepsilon) = 0$，$\text{cov}(y,\varepsilon) = 0$ 被破坏，我们已经知道会产生变量误差模型（Errors in Variables Model），实际上，这就是一个联立方程面临的问题。

第一节　概　述

对古典假设 $\text{cov}(x,\varepsilon) = 0$，$\text{cov}(y,\varepsilon) = 0$ 来说，y 是独立于 ε 的，x 也独立于 ε。但有时这条假设不容易满足。例如，在需求、供给模型中，需求模型为：$q = \alpha + \beta p + \varepsilon$，其中，$q$ 为需求量，p 为价格，ε 为扰动项或称作移动参数。

根据一般的经济学知识，我们知道供求变化有如下三种情况：

第一种情况，如图 12.1；在 ε 增大时，需求曲线将右移，q 与 p 同时增大。这是因为 $q = \alpha + \beta p + \varepsilon = (\alpha + \varepsilon) + \beta p$，当 ε 增大时，截距就扩大了，总需求曲线右移至 D'，q 增大到 q_1，p 增大到 p_1。

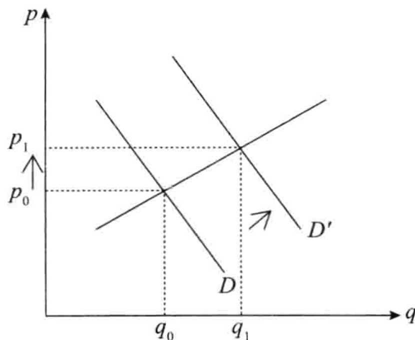


图 12.1

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



因此，此时 ε 与 p 存在一定关系，古典假设 $\operatorname{cov}(x,\varepsilon)=0,\operatorname{cov}(y,\varepsilon)=0$ 不再满足。

第二种情况，如图 12.2，此时供给曲线对价格完全没有反应，也即供给曲线价格弹性为零。在这种情况下，当 ε 增大时，仅有价格 p 的变化，而 q 没有任何变动。

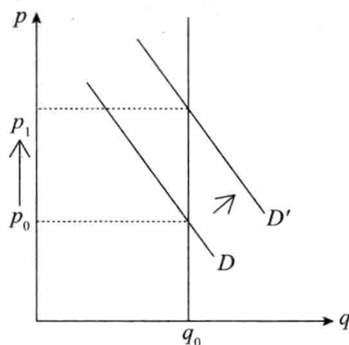


图 12.2

第三种情况，如图 12.3，此时供给曲线弹性无穷大，对价格反应极为敏感。在这种情况下，当 ε 增大，只能引起 q 的变化，而价格 p 不会有任何变化。

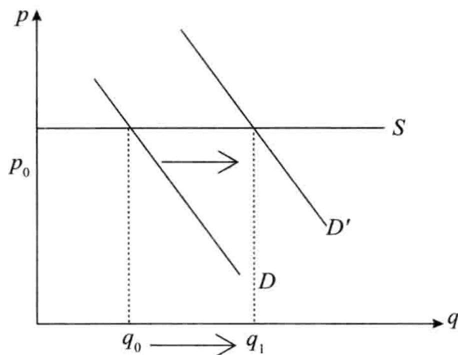


图 12.3

从上面的分析可知，在第一、第二种情况下，ε 变化不仅引起需求量 q 的变化，而且价格 p 也随之变化。这说明 ε 与 p 不独立，因此该问题不能简单地用一般的最小二乘法来解决。换言之，由于 $\operatorname{cov}(\varepsilon,p)\neq 0$，违反了古典假设，故而一般最小二乘法如被误用，就会产生一个有偏估计量。

由此可见，供给曲线的形状对于研究需求曲线来说有着十分重要的作用。研究需求函

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

数时，还应把供给函数考虑进去，对二者同时进行估计。这样的模型就是联立方程组模型(Simultaneous Equation Model)。

在联立方程组中存在着标准化问题。比如，在上面的例子中的需求方程可写作：$q = \alpha + \beta p + \varepsilon$，当然也可以写作：$p = \alpha' + \beta' q + u$。我们称 $q = \alpha + \beta p + \varepsilon$ 是根据 q 来标准化的方程，而 $p = \alpha' + \beta' q + u$ 是根据 p 标准化的方程。从理论上来说，上述两个方程应该是等价的，只不过是表达形式上的差异，因此估计方法无论是对以 p 标准化的方程，还是以 q 标准化的方程，估计的结果应是一致的。但事实上往往并非如此。对某些估计方法来说，不同的标准化形式，会得到不同的估计结果，这就是由标准化引起的问题。

例如，从 $q = \alpha + \beta p + \varepsilon$ 导出 $p = \alpha' + \beta' q + u$ 时，其前提条件是 $\beta \neq 0$。$\beta = 0$ 时就意味着需求无弹性，如图 12.4 所示。

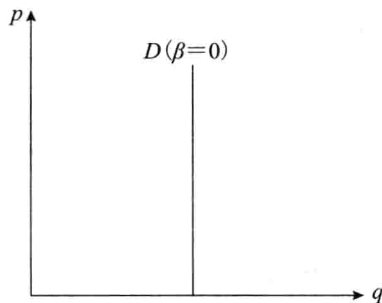


图 12.4

从 $p = \alpha' + \beta' q + u$ 转换到 $q = \alpha + \beta p + \varepsilon$ 时，其前提条件是 $\beta' \neq 0$。如果。$\beta' = 0$ 时意味着需求曲线弹性无穷大，如图 12.5 所示。

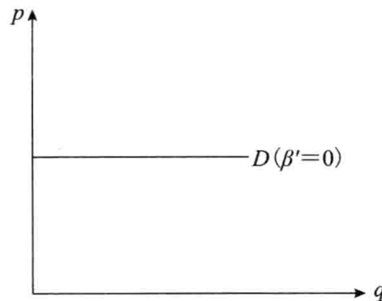


图 12.5

也就是说，若图 12.4、图 12.5 所示两种情况存在时，上述方程是不能相互转换的。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第二节　内生变量和外生变量

联立方程组中所有变量被分为两类：一类是内生变量（Endogenous Variables），也叫联合决定变量（Jointly - Determined Variables），这是由经济模型决定的变量；另一类是外生变量（Exogenous Variables），也叫前定变量（Predetermined Variables），是由外部因素决定的变量。

我们将以原油供给需求的简单模型为例，对联立方程组进行分析：

$$\begin{cases} \text{需求方程}: q = a_1 + b_1 p + c_1 y + \varepsilon_1 \\ \text{供给方程}: q = a_2 + b_2 p + c_2 R + \varepsilon_2 \end{cases}$$

其中，q 指原油总量，p 指原油价格，y 指国民收入，R 指降雨量。

很明显，q 、p 为内生变量，y 、R 为外生变量。当然，真正的原油供求模型并非如此简单，我们只不过想要把事情搞得简单些明了些，才设立上述方程的。

第三节　间接最小二乘法

间接最小二乘法是对付联立方程组模型的一种常用手段，其步骤如下：

步骤一　既然外生变量独立于 ε，我们可以分别用内生变量 p 、q 对外生变量 y 、R 进行回归。

步骤二　对原来的模型求解；

步骤三　把新求得的参数转化为原参数。

例如，原油供求模型为：

$$\begin{cases} q = a_1 + b_1 p + c_1 y + \varepsilon_1 & (12.1) \\ q = a_2 + b_2 p + c_2 R + \varepsilon_2 & (12.2) \end{cases}$$

第一步，用 q 对外生变量 y 、R 回归：

$$q = \pi_1 + \pi_2 y + \pi_3 R + v_1 \tag{12.3}$$

用 p 对外生变量 y 、R 回归：

$$p = \pi_4 + \pi_5 y + \pi_6 R + v_2 \tag{12.4}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



第二步，对原模型求解：

以 y 、R 两个外生变量为已知量，对式 12.1、12.2 联立方程求解，得出以 y 、R 表示的 q 、p 的解：

$$q = \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1} + \frac{c_1 b_2}{b_2 - b_1} y - \frac{c_2 b_1}{b_2 - b_1} R + \text{误差项} \tag{12.5}$$

$$p = \frac{a_2 - a_1}{b_2 - b_1} + \frac{c_1}{b_2 - b_1} y - \frac{c_2}{b_2 - b_1} R + \text{误差项} \tag{12.6}$$

第三步，转化成原参数：

将式 12.5、12.6 与第一步得出的式 12.3、12.4 分别对照，可得：

$$\pi_1 = \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1}$$

$$\pi_2 = \frac{c_1 b_2}{b_2 - b_1}$$

$$\pi_3 = -\frac{c_2 b_1}{b_2 - b_1}$$

$$\pi_4 = \frac{a_2 - a_1}{b_2 - b_1}$$

$$\pi_5 = \frac{c_1}{b_2 - b_1}$$

$$\pi_6 = -\frac{c_2}{b_2 - b_1}$$

解方程，有：

$$\hat{b}_1 = \frac{\hat{\pi}_3}{\hat{\pi}_6}$$

$$\hat{b}_2 = \frac{\hat{\pi}_2}{\hat{\pi}_5}$$

$$\hat{c}_1 = -\hat{\pi}_5(\hat{b}_1 - \hat{b}_2)$$

$$\hat{c}_2 = \hat{\pi}_6(\hat{b}_1 - \hat{b}_2)$$

$$\hat{a}_1 = \hat{\pi}_1 - \hat{b}_1 \hat{\pi}_4$$

$$\hat{a}_2 = \hat{\pi}_1 - \hat{b}_2 \hat{\pi}_4$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

一般来说，我们称模型：

$$\begin{cases} q = a_1 + b_1 p + c_1 y + \varepsilon_1 \\ q = a_2 + b_2 p + c_2 R + \varepsilon_2 \end{cases}$$

为结构方程（Structure Form），对内生变量的位置不一定要求，可在等号的左边，也可在等号的右边，如方程组中内生变量 q 或 p 可在等号两端。我们又称模型：

$$\begin{cases} q = \dfrac{a_1 b_2 - a_2 b_1}{b_2 - b_1} + \dfrac{c_1 b_2}{b_2 - b_1} y - \dfrac{c_2 b_1}{b_2 - b_1} R + \text{误差项} \\ p = \dfrac{a_2 - a_1}{b_2 - b_1} + \dfrac{c_1}{b_2 - b_1} y - \dfrac{c_2}{b_2 - b_1} R + \text{误差项} \end{cases}$$

为约简方程（Reduced Form）。约简方程要求内生变量必须在等号左边，而外生变量必须在等号右边。因此，模型

$$\begin{cases} q = \pi_1 + \pi_2 y + \pi_3 R + \text{误差项} \\ p = \pi_4 + \pi_5 y + \pi_6 R + \text{误差项} \end{cases}$$

也是约简方程，π_i 为约简方程参数，而 a_i，b_i，c_i 则为结构方程参数。

第四节　识别问题

我们可能已经注意到了，在间接最小二乘法第三步中，并不总是可以用约简方程参数得到原结构方程参数，有时可能会得到多个结构方程系数，有时则可能一个也得不到。这就是联立方程组的识别问题（Identification Problem）。

比如：

$$\begin{cases} \text{需求方程：} q = a_1 + b_1 p + c_1 y + \varepsilon_1 \\ \text{供给方程：} q = a_2 + b_2 p + \varepsilon_2 \end{cases}$$

这一方程组的约简方程为：

$$\begin{cases} q = \dfrac{a_1 b_2 - a_2 b_1}{b_2 - b_1} + \dfrac{c_1 b_2}{b_2 - b_1} y + \text{误差项} \\ p = \dfrac{a_2 - a_1}{b_2 - b_1} + \dfrac{c_1}{b_2 - b_1} y + \text{误差项} \end{cases}$$

在这类情况下，只能得到：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\hat{b}_2 = \frac{\hat{\pi}_2}{\hat{\pi}_4}$$

$$\hat{a}_2 = \hat{\pi}_1 - \hat{b}_2\hat{\pi}_3$$

却不能得到会 $\hat{a}_1$ 、$\hat{b}_1$ 、$\hat{c}_1$ 这三个结构参数。

这说明供给方程 $q = a_2 + b_2p + \varepsilon_2$ 是可识别的，因为可以得到它要求的全部系数，而需求方程 $q = a_1 + b_1p + c_1y + \varepsilon_1$ 则是不可识别的。

再如：

$$\begin{cases} 需求方程: q = a_1 + b_1p + c_1y + d_1R + \varepsilon_1 \\ 供给方程: q = a_2 + b_2p + \varepsilon_2 \end{cases}$$

其约简方程是

$$\begin{cases} q = \dfrac{a_1b_2 - a_2b_1}{b_2 - b_1} + \dfrac{c_1b_2}{b_2 - b_1}y + \dfrac{d_1b_2}{b_2 - b_1}R + 误差项 \\ p = \dfrac{a_2 - a_1}{b_2 - b_1} + \dfrac{c_1}{b_2 - b_1}y + \dfrac{d_1}{b_2 - b_1}R + 误差项 \end{cases}$$

这时，能得到 $\hat{b}_2 = \dfrac{\hat{\pi}_2}{\hat{\pi}_5}$ 或 $\hat{b}_2 = \dfrac{\hat{\pi}_3}{\hat{\pi}_6}$，因此 $\hat{a}_2 = \hat{\pi}_1 - \hat{b}_2$ 也就必然有两个值。但仍不能得到 $\hat{a}_1$ 、$\hat{b}_1$ 、$\hat{c}_1$ 、$\hat{d}_1$ 的估计量。

我们看到在供给方程 $q = a_2 + b_2p + \varepsilon_2$ 的每个系数都有 1 个以上的解，这种情况称为过度识别；而需求方程 $q = a_1 + b_1p + c_1y + d_1R + \varepsilon_1$ 的每一个系数都没有解，这种情况称为不可识别。严格地说，如果能够从约简方程得到结构方程参数的唯一解，那么，这个方程就是完全的识别（Exactly - Identified）；如果得到多个结构方程参数的估计量，称为过度识别（Over - Identified）；如果得不到这个结构方程参数的估计量，称为不可识别（Under - Identified）。值得注意的是，识别是针对某个方程，而不是针对整个方程组而言的。

对于识别的判断有一个必要条件，称为识别的阶的条件（Order Condition），这个条件规则如下：

1. $k = g - 1$，方程完全可识别；
2. $k > g - 1$，方程过度识别；
3. $k < g - 1$，方程不可识别。

其中，g 指该方程所包含的内生变量个数，k 指该方程没包含的系统中所有变量个数，这包括内生、外生变量。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



例如

$$需求方程：q = a_1 + b_1 p + c_1 y + d_1 R + \varepsilon_1 \tag{12.7}$$

$$供给方程：q = a_2 + b_2 p + \varepsilon_2 \tag{12.8}$$

对方程 12.8 而言：$g = 2$，内生变量为 (q,p)。

而方程 12.8 以外的方程只有方程 12.7，在方程 12.8 中没有出现而在方程 12.7 中出现的变量个数为 2 个，(y,R)，即：$k = 2$。因此，$k > g - 1$，方程 12.8 将过度识别。详解见表 12－1。

表 12－1

项目 / 方程	g	k	$k-(g-1)>0$？
供给方程	2	2	2－1＋1＞0，过度识别
需求方程	2	0	0－2＋1＜0，不可识别

这样，我们根本不用解这个方程组，就可以知道哪些可以识别，哪些不可识别。唯一要注意的是，这个规则仅是方程识别的必要条件，其充要条件我们将在下一节讨论。

第五节　识别的充要条件

在讨论这个充要条件时，我们仍以如下供求模型为例：

$$\begin{cases} 需求方程：q = a_1 + b_1 p + c_1 y + d_1 R + \varepsilon_1 \\ 供给方程：q = a_2 + b_2 p + \varepsilon_2 \end{cases}$$

将内生变量、外生变量分布以表 12－2 表示：

表 12－2

	内生变量		外生变量	
	q	p	y	R
需求方程	X	X	X	X
供给方程	X	X	O	O

其中，X 表示该变量在该方程中存在，O 表示该变量在该方程中不存在。

熟悉上表后，我们列出步骤：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

1. 考察哪一行（即考察哪一方程）就删掉哪一行。

2. 把这一行中元素为 O 的哪些列排出来，组成新的行列式。

3. 如果在挑出的行列式中有 $g-1$ 行不全为 O，那么方程就是可以完全识别。否则就是识别不了（或许是过度识别，或许是不可识别）。注意，这里 g 是指全部内生变量的个数。

下面是更为复杂的情况，我们可以看出具体的操作步骤：

表 12－3

	内生变量			外生变量		
方程	y_1	y_2	y_3	z_1	z_2	z_3
1	X	O	X	X	O	X
2	X	O	O	X	O	X
3	O	X	X	X	X	O

例如对方程 1 来说：

第一步，删去第一行（因为就是考察第一个方程）。

表 12－4

	内生变量			外生变量		
方程	y_1	y_2	y_3	z_1	z_2	z_3
1	~~X~~	~~O~~	~~X~~	~~X~~	~~O~~	~~X~~
2	X	O	O	X	O	X
3	O	X	X	X	X	O

第二步，找出第一行中原为 O 的列重组。

表 12－5

	y_1	y_2	y_3	z_1	z_2	z_3			
1	~~X~~	~~O~~	~~X~~	~~X~~	~~O~~	~~X~~		y_2	z_2
2	X	O	O	X	O	X	⇨	O	O
3	O	X	X	X	X	O		X	X

第三步，因为内生变量总量为 $g=3$，那么 $g-1=2$，应该有两行不完全为 O，但是重组后的表中只有一行不完全为 O，所以该方程不可识别。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

我们可以依据此方程来检验方程 2、方程 3 是否可以识别。

第六节 估计方法

由于存在 $\text{cov}(x_i,\varepsilon_i)\neq 0$ 违背古典假设的情况，如果再采用普通最小二乘法来解决已不能得出满意的结论。正如我们在以前碰到的异方差、自回归问题一样，如果我们贸然地使用普通最小二乘法，则可能产生许多不良后果。那么，此时该如何进行估计呢？联立方程组的估计方法有两种：一种方法是单个方程估计方法（Single Equation Estimation），又称有限信息方法（Limited Information Method）；另一种为系统方程估计方法（System Equation Estimation），又称完全信息方法（Full Information Method）。我们这里主要讨论前一种估计方法。在单个方程估计方法中，我们分别估计每一个方程，并只对单个方程的系数及其约束感兴趣。在估计单个方程时，我们经常使用工具变量法。

一、工具变量法

工具变量方法，又称 IV 法，它是对单个方程估计时经常用的一种方法。所谓工具变量，就是指与方程的误差项无关，与解释变量有关的变量。

例如，对模型 $y=\beta x+\varepsilon,\text{cov}(x,\varepsilon)\neq 0$ 而言，这是个典型误差变量模型，是不能使用普通最小二乘法的。但是我们能否找到一个变量 z 来代替 x，并且变量 z 具有如下性质：$\text{cov}(x,z)\neq 0,\text{cov}(z,\varepsilon)=0$。如果能够找到，我们就可以得到 β 的一致估计量，即当 $n\to\infty$ 时，该估计量与 β 相差很小。

由于工具变量 z 满足 $\text{cov}(z,\varepsilon)=0$。因此

$$\text{cov}(z,\varepsilon)=0\Rightarrow\frac{1}{n}\sum(z-\bar{z})(\varepsilon-\bar{\varepsilon})=0$$

$$\Rightarrow\frac{1}{n}\sum z\varepsilon=0\Rightarrow\frac{1}{n}\sum z(y-\beta x)=0\Rightarrow\hat{\beta}_{IV}=\frac{\sum zy}{\sum zx}$$

我们自然记得用普通最小二乘法得到的 $\hat{\beta}_{LS}=\dfrac{\sum xy}{\sum x^2}$。$\hat{\beta}_{IV}$ 与 $\hat{\beta}_{LS}$ 之间的差别只在一个 z 上，但意义却相差很大。工具变量法是根据工具变量具有 $\text{cov}(z,\varepsilon)=0$ 性质得到 $\hat{\beta}_{IV}=$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



$\frac{\sum zy}{\sum zx}$；而普通最小二乘法是从 Min$\sum(y-\beta x)^2$ 入手，获得 $\hat{\beta}_{LS}=\frac{\sum xy}{\sum x^2}$。

下面，我们来仔细考察 $\hat{\beta}_{IV}$ 的有关问题：

$$\hat{\beta}_{IV}=\frac{\sum zy}{\sum zx}=\frac{\sum z(\beta x+\varepsilon)}{\sum zx}$$

$$=\beta+\frac{\sum z\varepsilon}{\sum zx}=\beta+\frac{\frac{1}{n}\sum z\varepsilon}{\frac{1}{n}\sum zx}$$

当 $n\to\infty$时，$\frac{\frac{1}{n}\sum z\varepsilon}{\frac{1}{n}\sum zx}=\frac{\mathrm{cov}(z,\varepsilon)}{\mathrm{cov}(z,x)}$。

由于工具变量 z 满足 $\mathrm{cov}(z,\varepsilon)=0$，$\mathrm{cov}(z,x)\neq 0$，因此，此时的 $\mathrm{plim}\hat{\beta}_{IV}=\beta$，也即使用工具变量法得到的估计量 $\hat{\beta}_{IV}$ 为 β 的一致估计量。

工具变量法给我们带来如此好处，那么如何去寻找工具变量呢？下面对如下模型进行研究考察：

$$\begin{cases} y_1=b_1y_2+c_1z_1+d_1z_2+\varepsilon_1 & (12.9)\\ y_2=b_2y_1+c_2z_3+\varepsilon_2 & (12.10)\end{cases}$$

其中，y_1，y_2 为内生变量；z_1，z_2，z_3 为外生变量。

先看方程 12.9：z_1 是外生变量，与 ε_1 无关，但是 z_1 与 y_2 存在很显著的关系，因此毫无疑问 z_1 可以做工具变量。同样可知，z_2 也能够做工具变量。z_3 可不可以呢？z_3 是外生变量，与 ε_1 无关，且 z_3 可以影响 y_2，因此 z_3 满足 $\mathrm{cov}(z_3,y_2)\neq 0$，$\mathrm{cov}(z_3,\varepsilon_1)=0$，$z_3$ 也可以做工具变量。

根据工具变量的定义，有：

$$\begin{cases}\mathrm{cov}(z_1,\varepsilon_1)=0\\ \mathrm{cov}(z_2,\varepsilon_1)=0\\ \mathrm{cov}(z_3,\varepsilon_1)=0\end{cases}\Rightarrow\begin{cases}\frac{1}{n}\sum z_1\varepsilon_1=0\\ \frac{1}{n}\sum z_2\varepsilon_1=0\\ \frac{1}{n}\sum z_3\varepsilon_1=0\end{cases}\Rightarrow\begin{cases}\frac{1}{n}\sum z_1(y_1-b_1y_2-c_1z_1-d_1z_2)=0 & (12.11)\\ \frac{1}{n}\sum z_2(y_1-b_1y_2-c_1z_1-d_1z_2)=0 & (12.12)\\ \frac{1}{n}\sum z_3(y_1-b_1y_2-c_1z_1-d_1z_2)=0 & (12.13)\end{cases}$$

对比普通最小二乘法，可以发现工具变量法有如下方程成立：

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\begin{cases} \sum z_1\varepsilon_1 = 0 \\ \sum z_2\varepsilon_1 = 0 \\ \sum y_2\varepsilon_1 = 0 \end{cases} \Rightarrow \begin{cases} \sum z_1(y_1 - b_2y_2 - c_1z_1 - d_1z_2) = 0 & (12.14) \\ \sum z_2(y_1 - b_2y_2 - c_1z_1 - d_1z_2) = 0 & (12.15) \\ \sum y_2(y_1 - b_2y_2 - c_1z_1 - d_1z_2) = 0 & (12.16) \end{cases}$$

通过对这两个方程组进行对比可知：只有式 12.13、12.16 两个方程不同。使用工具变量法可得到 b_1、c_1、d_1 的估计量，而普通最小二乘法也可得到 b_1、c_1、d_1 的估计量，但这两种估计量有所区别。

对方程 12.10 来说，只有 b_2、c_2 两个未知参数，却有 z_1、z_2、z_3 三个变量来做工具变量。

从 $\begin{cases} \text{cov}(z_3,\varepsilon_2) = 0 \\ \text{cov}(z_1,\varepsilon_2) = 0 \end{cases}$，可得到 $\hat{c}_2$，$\hat{b}_2$，而从 $\begin{cases} \text{cov}(z_3,\varepsilon_2) = 0 \\ \text{cov}(z_2,\varepsilon_2) = 0 \end{cases}$下，同样也可得到不同的 $\hat{c}_2$，$\hat{b}_2$。

前面的分析方法告诉我们，该方程为过度识别。此时我们有两组 b_2、c_2 的值，究竟选哪一组呢？实际上我们为了慎重起见，使用工具变量的加权值来解决这一问题。我们后面会进一步讨论这个问题。方程组识别的阶条件与我们是否有足够的外生变量来作为工具变量问题有关。如果方程不可识别，我们就没有足够的外生变量来做工具变量。如果方程是完全识别，我们正好有足够的外生变量来充当工具变量；如果过度识别，则实际的外生变量就比我们所需要的工具变量要多得多。

现在，回过头来看方程 12.9，此时我们以 $\hat{y}_2$ 作为工具变量。

第一步，以所有外生变量对 y_1，y_2 进行回归，得到：

$$\hat{y}_1 = \hat{a}_{11}z_1 + \hat{a}_{12}z_2 + \hat{a}_{13}z_3$$

$$\hat{y}_2 = \hat{a}_{21}z_1 + \hat{a}_{22}z_2 + \hat{a}_{23}z_3$$

第二步，使用 $\hat{y}_2$、z_1、z_2 作为工具变量。这里我们已知外生变量一定能够作为工具变量，而 $\hat{y}_2$ 呢？我们会证明它也可以成为工具变量：

$$\begin{cases} \text{cov}(\hat{y}_2,\varepsilon_1) = 0 \\ \text{cov}(z_1,\varepsilon_1) = 0 \\ \text{cov}(z_2,\varepsilon_1) = 0 \end{cases} \Rightarrow \begin{cases} \sum \hat{y}_2(y_1 - b_1y_2 - c_1z_1 - d_1z_2) = 0 \\ \sum z_1(y_1 - b_1y_2 - c_1z_1 - d_1z_2) = 0 \\ \sum z_1(y_1 - b_1y_2 - c_1z_1 - d_1z_2) = 0 \end{cases}$$

单看

$$\sum \hat{y}_2(y_1 - b_1y_2 - c_1z_1 - d_1z_2) = 0$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



$$\Rightarrow \sum \hat{y}_2 \varepsilon_1 = 0$$

$$\Rightarrow \sum (\hat{a}_{21} z_1 + \hat{a}_{22} z_2 + \hat{a}_{23} z_3) \varepsilon_1 = 0$$

$$\Rightarrow \hat{a}_{21} \sum z_1 \varepsilon_1 + \hat{a}_{22} \sum z_2 \varepsilon_1 + \hat{a}_{23} \sum z_3 \varepsilon_1 = 0$$

由于 $\sum z_1 \varepsilon_1 = 0$，$\sum z_2 \varepsilon_1 = 0$ 成立，则有：$\sum z_3 \varepsilon_1 = 0$

也就是说，我们可以从 $\sum \hat{y}_2 (y_1 - b_1 y_2 - c_1 z_1 - d_1 z_2) = 0$ 得到 $\sum z_3 \varepsilon_1 = 0$，那么上面的方程组也就等价于

$$\begin{cases} \sum z_3 \varepsilon_1 = 0 \\ \sum z_1 \varepsilon_1 = 0 \\ \sum z_2 \varepsilon_1 = 0 \end{cases} \quad 即 \begin{cases} \operatorname{cov}(z_3, \varepsilon_1) = 0 \\ \operatorname{cov}(z_2, \varepsilon_1) = 0 \\ \operatorname{cov}(z_1, \varepsilon_1) = 0 \end{cases}$$

我们由此得出结论：用 $\hat{y}_2$、z_1、z_2 作工具变量与用 z_1、z_2、z_3 作工具变量是等价的。因为 y_2 是与 ε_1 相关的，所以要找一个变量代替 y_2，而这个变量与 ε_1 无关，仅与 y_2 有关。因此，我们可以找 z_3 代替 y_2，也可以找 y_2 的拟合值 $\hat{y}_2$ 作工具变量。

再考察一下方程 12.10。此时使用 $\hat{y}_1$ 作工具变量，那么有

$$\begin{cases} \operatorname{cov}(\hat{y}_1, \varepsilon_2) = 0 \\ \operatorname{cov}(z_3, \varepsilon_2) = 0 \end{cases} \Rightarrow \begin{cases} \sum \hat{y}_1 (y_2 - b_2 y_1 - c_2 z_3) = 0 \\ \sum z_3 (y_2 - b_2 y_1 - c_2 z_3) = 0 \end{cases}$$

$$\sum \hat{y}_1 (y_2 - b_2 y_1 - c_2 z_3) = 0$$

$$\Rightarrow \sum (a_{11} z_1 + a_{12} z_2 + a_{13} z_3) \varepsilon_2 = 0$$

$$\Rightarrow a_{11} \sum z_1 \varepsilon_2 + a_{12} \sum z_2 \varepsilon_2 + a_{13} \sum z_3 \varepsilon_2 = 0$$

$$\Rightarrow a_{11} \sum z_1 \varepsilon_2 + a_{12} \sum z_2 \varepsilon_2 = 0 \ (因为 \sum z_3 \varepsilon_2 = 0)$$

这就是工具变量的加权值组合，其中 a_{11}，a_{12} 为权数。

综上所述，我们在找工具变量时，一般有两种方法：一是找本方程以外的外生变量作工具变量；二是找内生变量的拟合值作工具变量。对这两种方法而言，如果方程为完全识别，则两种方法得到的估计值是等价的。否则，二者可能会有所差异。也就是说，当我们对模型 $y = \beta x + \varepsilon$，$\operatorname{cov}(z, \varepsilon) \neq 0$ 进行分析时，工具变量法就是让我们跨过 $\operatorname{cov}(x, \varepsilon) \neq 0$ 另找一个变量 z，满足 $\operatorname{cov}(z, x) \neq 0$，$\operatorname{cov}(z, \varepsilon) = 0$，并以此来代替 x。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

二、两阶段最小二乘法

两阶段最小二乘法，又记作 TSLS 法，由瑟耳（G. Theil）在 1958 年提出。其具体步骤如下：

1. 用最小二乘法估计约简方程，并得到 y_i 的估计值 $\hat{y}_i$ 。

2. 用 $\hat{y}_i$ 各自代替等式右侧内生变量并用最小二乘法进行回归。

TSLS 法与 IV 法不同之处在于把 $\hat{y}_i$ 作为解释变量，而不是作为工具变量。但这两种方法的估计结果是一样的。

考虑 $y_1 = b_1 y_2 + c_1 z_1 + \varepsilon_1$，其中，$y_1$、$y_2$ 为内生变量；z_1、z_2、z_3、z_4 为外生变量。

TSLS 法首先以 y_2 对 z_1、z_2、z_3、z_4 进行回归，得到一个 $\hat{y}_2$，并且自然有 $y_2 = \hat{y}_2 + v_2$，然后对 $y_1 = b_2 \hat{y}_2 + c_1 z_1 + \varepsilon_1^*$ 进行回归，此时已用 $\hat{y}_2$ 代替了 y_2，有

$$\begin{cases} \sum \hat{y}_2 \varepsilon_1^* = 0 \\ \sum z_1 \varepsilon_1^* = 0 \end{cases} \Rightarrow \begin{cases} \sum \hat{y}_2 (y_1 - b_1 \hat{y}_2 - c_1 z_1) = 0 \\ \sum z_1 (y_1 - b_1 \hat{y}_2 - c_1 z_1) = 0 \end{cases} \tag{12.17}$$

我们会注意到，在工具变量法下，如用 $\hat{y}_2$ 作为工具变量，则有：

$$\begin{cases} \mathrm{cov}(\hat{y}_2, \varepsilon_1) = 0 \\ \mathrm{cov}(z_1, \varepsilon_1) = 0 \end{cases} \Rightarrow \begin{cases} \sum \hat{y}_2 (y_1 - b_2 y_2 - c_1 z_1) = 0 \\ \sum z_1 (y_1 - b_2 y_2 - c_1 z_1) = 0 \end{cases} \tag{12.18}$$

可见方程组 12.17 与 12.18 的唯一差别就在于，是 $y_1 - b_1 \hat{y}_2 - c_1 z_1$，还是 $y_1 - b_2 y_2 - c_1 z_1$ 。

由于 $y_2 = \hat{y}_2 + v_2$，代入到方程组 Ⅱ 中，可得：

$$\begin{cases} \sum \hat{y}_2 (y_1 - b_1 \hat{y}_2 - c_1 z_1) - b_1 \sum \hat{y}_2 v_2 = 0 \\ \sum z_1 (y_1 - b_1 \hat{y}_2 - b_1 z_1) - c_1 \sum z_1 v_2 = 0 \end{cases}$$

而 $\sum \hat{y}_2 v_2 = 0$ 并且 $\sum z_1 v_2 = 0$，所以两方程是等价的。

我们还注意到，在 TSLS 第二步中，我们以拟合值 $\hat{y}_i$ 代替右侧的内生变量 y_i 能否以拟合值 $\hat{y}_i$ 代替全部内生变量 y_i 呢？

对 $y_1 = b_1 y_2 + c_1 z_1 + \varepsilon_1$ 来说，代替所有内生变量就意味着有 $\hat{y}_1 = b_1 \hat{y}_2 + c_1 z_1 + \omega$，于是使用最小二乘法可得：

$$\begin{cases} \sum \hat{y}_2 (\hat{y}_1 - b_2 \hat{y}_2 - c_1 z_1) = 0 \\ \sum z_1 (\hat{y}_1 - b_2 \hat{y}_2 - c_1 z_1) = 0 \end{cases}$$

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

而 TSLS 只代替右边的内生变量，有：

$$\begin{cases}\sum \hat{y}_2(y_1 - b_1\hat{y}_2 - c_1z_1) = 0 \\ \sum z_1(y_1 - b_1\hat{y}_2 - c_1z_1) = 0\end{cases}$$

仍旧以 $y_1 = \hat{y}_1 + v_1$ 代入上式，会有：

$$\begin{cases}\sum \hat{y}_2(\hat{y}_1 + v_1 - b_1\hat{y}_2 - c_1z_1) = 0 \\ \sum z_1(\hat{y}_1 + v_1 - b_1\hat{y}_2 - c_1z_1) = 0\end{cases} \Rightarrow \begin{cases}\sum \hat{y}_2(\hat{y}_1 - b_1\hat{y}_2 - c_1z_1) + \sum \hat{y}_2v_1 = 0 \\ \sum z_1(\hat{y}_1 - b_1\hat{y}_2 - c_1z_1) + \sum z_1v_1 = 0\end{cases}$$

由于 z_1 为外生变量，与 v_1 无关，有 $\sum z_1v_1 = 0$，而 $\sum \hat{y}_2v_1 = 0$ 也是成立的。因此，两个方程组仍是等价的。这意味着我们把 TSLS 扩展了，可以只用拟合值 $\hat{y}_i$ 代替右边所有内生变量 y_i，也可以用拟合值 $\hat{y}_i$ 代替所有内生变量 y_i 。

当然，在 TSLS 方法中由于以 $\hat{y}_2$ 代替 y_2，$\hat{y}_1$ 代替 y_1，会使第二步中得到的估计标准残差有偏。

考虑模型 $y_1 = \beta y_2 + \varepsilon$，其中 y_1 、y_2 为内生变量。

首先，我们以 TSLS 法估计约简方程，得到 $\hat{y}_2$，并且有 $y_2 = \hat{y}_2 + v_2$（其中 $\hat{y}_2$ 为 y_2 对所有外生变量回归得到的拟合值）。

其次，（1）在工具变量法下，使用 $\hat{y}_2$ 为工具变量，那么有：

$$\operatorname{cov}(\hat{y}_2, \varepsilon_i) = 0 \Rightarrow \sum \hat{y}_2(y_1 - \beta y_2) = 0 \Rightarrow \hat{\beta}_{IV} = \frac{\sum \hat{y}_2y_1}{\sum \hat{y}_2y_2}$$

（2）在 TSLS 法下，对 $y_1 = \beta\hat{y}_2 + \varepsilon$ 进行回归，那么有：

$$\sum \hat{y}_2(y_1 - \beta\hat{y}_2) = 0 \Rightarrow \hat{\beta}_{TSLS} = \frac{\sum \hat{y}_2y_1}{\sum \hat{y}_2^2}$$

$\hat{\beta}_{IV}$ 与 $\hat{\beta}_{TSLS}$ 二者仅仅分母有一字母之差，下面我们可以证明二者是等价的。

$$\begin{aligned}\sum \hat{y}_2y_2 &= \sum \hat{y}_2(\hat{y}_2 + v_2) \\ &= \sum \hat{y}_2^2 + \sum \hat{y}_2v_2 \quad (\sum \hat{y}_2v_2 = 0) \\ &= \sum \hat{y}_2^2\end{aligned}$$

那么：$\hat{\beta} = \hat{\beta}_{IV} = \hat{\beta}_{TSLS} = \dfrac{\sum \hat{y}_2y_1}{\sum \hat{y}_2^2}$

而

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

$$\hat{\beta}=\frac{\sum\hat{y}_2y_1}{\sum\hat{y}_2^2}=\frac{\sum\hat{y}_2(\beta y_2+\varepsilon)}{\sum\hat{y}_2^2}=\frac{\beta\sum\hat{y}_2y_2+\sum\hat{y}_2\varepsilon}{\sum\hat{y}_2^2}$$

$$=\beta\frac{\sum\hat{y}_2y_2}{\sum\hat{y}_2^2}+\frac{\sum\hat{y}_2\varepsilon}{\sum\hat{y}_2^2}=\beta+\frac{\frac{1}{n}\sum\hat{y}_2\varepsilon}{\frac{1}{n}\sum\hat{y}_2^2}$$

当 $n\to\infty$，$\frac{\frac{1}{n}\sum\hat{y}_2\varepsilon}{\frac{1}{n}\sum\hat{y}_2^2}=\frac{\text{cov}(\hat{y}_2,\varepsilon)}{Var(\hat{y}_2)}$。由于 $\text{cov}(\hat{y}_2,\varepsilon)=0$，因此，$\text{plim}\hat{\beta}=\beta$。

这说明无论是工具变量法还是 TSLS 法，都可以得到 β 的一致估计量。

再看，$Var(\hat{\beta})=Var(\beta+\frac{\sum\hat{y}_2\varepsilon}{\sum\hat{y}_2^2})=\frac{\sum\hat{y}_2^2Var(\varepsilon)}{(\sum\hat{y}_2^2)^2}=\frac{\sigma_\varepsilon^2}{\sum\hat{y}_2^2}$。这是 β 的真正的方差。其中，σ_ε^2 表示 ε 的方差。

在计算机中直接进行二阶段最小二乘法计算时，我们用的是

$$\begin{aligned}y_1&=\beta y_2+\varepsilon=\beta(\hat{y}_2+v_1)+\varepsilon\\&=\beta\hat{y}_2+(\beta v_1+\varepsilon)\\&=\beta\hat{y}_2+\omega\quad(\omega=\beta v_1+\varepsilon)\end{aligned}$$

所以用 y_1 对 $\hat{y}_2$ 进行回归时，得到的残差为 ω，而不是 ε，此时 β 的标准差为 $\sqrt{\frac{\sigma_\omega^2}{\sum\hat{y}_2^2}}$，这是个计算机给出未经调整的 $\hat{\beta}_{LS}$ 标准差。怎样把这个计算机结果调整成真正的方差呢？我们可以乘以 $\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\omega^2}}$ 进行调整。此时 $\hat{\sigma}_\varepsilon^2=\frac{1}{n}\sum(y_1-\hat{\beta}y_2)^2$，$\hat{\sigma}_\omega^2=\frac{1}{n}\sum(y_1-\hat{\beta}\hat{y}_2)^2$。实际上计算机在进行 TSLS 估计时，会进行相应的调整。

在 EViews 软件包中只要我们给出如下命令 TSLS 即可：

TSLS y_1 y_2 @ z_1 z_2 z_3

这里 y_2 为内生变量，z_1，z_2，z_3 为工具变量。

三、有限信息最大似然法

有限信息最大似然方法简称 LIML，由 T. W. Anderson 和 Rubin 于 1949 年提出。在 1958 年瑟耳推出 TSLS 方法以前，是最为流行的估计方法。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

考虑模型：$y_1 = b_1y_2 + c_1z_1 + d_1z_2 + \varepsilon$，其中，$y_1$、$y_2$、$y_3$ 为内生变量，z_1、z_2、z_3 为外生变量。

LIML 的具体步骤如下：

1. 任找一个 b_1，计算出 $y_1^* = y_1 - b_1y_2$。
2. 对 $y_1^* = a_1z_1 + a_1z_2$ 进行回归，得到 ESS，记作 ESS_1。
3. 对 $y_1^* = a_1^*z_1 + a_2^*z_2 + a_3^*z_3$ 进行回归，得到另一个 ESS，记作 ESS_2。
4. 因为 z_3 与决定 y_1^* 无关，因此 LIML 方法就是通过 $\text{Min}(\frac{ESS_1 - ESS_2}{ESS_2})$ 或 $\text{Min}(\frac{ESS_1}{ESS_2})$ 来选到 b_1。
5. b_1 被确定后，就可以通过对 $y_1^* = a_1^{**}z_1 + a_2^{**}z_2 + \varepsilon$ 进行回归，得到 c_1，d_1 的估计量。

对比 LIML 和 TSLS，会发现它们之间有一定关系：

1. TSLS 法下是使 $ESS_1 - ESS_2$ 最小，$\text{Min}(ESS_1 - ESS_2)$，而 LIML 下却要使 $\frac{ESS_1}{ESS_2}$ 最小。
2. 如果方程为完全识别，那么 TSLS 与 LIML 的估计量是一样的。

四、估计方法与标准化问题

目前为止，本章里我们已讨论过四种估计方法，即 FIML 方法、LIML 方法、IV 方法和 TSLS 方法。我们知道，不同的估计方法对标准化要求有所不同。上述四种方法对标准化的要求见表 12－5。

表 12－5

估计方法	不依赖标准化	依赖标准化
FIML	√	
LIML	√	
IV 法	完全识别时成立	过度识别时成立
TSLS	完全识别时成立	过度识别时成立

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第七节 运用最小二乘法估计联立方程组问题

考虑供求模型：

$$\begin{cases} q_t = \beta p_t + \varepsilon_t & \text{需求方程 I} \\ q_t = \alpha p_t + v_t & \text{供给方程 II} \end{cases}$$

$$\beta < 0, Var(\varepsilon_t) = \sigma_t^2$$

$$\text{且}, \alpha > 0, Var(v_t) = \sigma_v^2 \text{。}$$

$$\operatorname{cov}(\varepsilon_t, v_t) = \sigma_{tv}$$

在实际中，如果我们直接用销售 q_t 与价格水平 p_t 进行回归，将无法判断所得到的是供给曲线还是需求曲线，因为方程组 I 、II 都是不可识别的。

采用最小二乘法，可得：

$$\hat{\beta}_{LS} = \frac{\sum p_t q_t}{\sum p_t^2} = \frac{\sum p_t(\beta p_t + \varepsilon_t)}{\sum p_t^2} = \beta + \frac{\sum p_t \varepsilon_t}{\sum p_t^2}$$

当 $n \to \infty$, $\dfrac{\sum p_t\varepsilon_t}{\sum p_t^2} = \dfrac{\operatorname{cov}(p_t,\varepsilon_t)}{Var(p_t)}$, 则 $\operatorname{plim}\hat{\beta}_{LS} = \beta + \dfrac{\operatorname{cov}(p_t,\varepsilon_t)}{Var(p_t)}$ 。

根据供给等于需求这一均衡条件，可知：

$$\beta p_t + \varepsilon_t = \alpha p_t + v_t$$

那么

$$p_t = \frac{v_t - \varepsilon_t}{\beta - \alpha}$$

$$\operatorname{cov}(p_t, \varepsilon_t) = \operatorname{cov}\left(\frac{v_t - \varepsilon_t}{\beta - \alpha}, \varepsilon_t\right) = \frac{1}{\beta - \alpha}\operatorname{cov}(v_t - \varepsilon_t, \varepsilon_t) = \frac{\sigma_{\varepsilon v} - \sigma_\varepsilon^2}{\beta - \alpha}$$

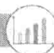
$$Var(p_t) = Var\left(\frac{v_t - \varepsilon_t}{\beta - \alpha}\right) = \frac{\sigma_v^2 + \sigma_\varepsilon^2 - 2\sigma_{\varepsilon v}}{(\beta - \alpha)^2}$$

当 $n \to \infty$时，

$$\operatorname{plim}\hat{\beta}_{LS} = \beta + \frac{\dfrac{\sigma_{\varepsilon v} - \sigma_\varepsilon^2}{\beta - \alpha}}{\dfrac{\sigma_v^2 + \sigma_\varepsilon^2 - 2\sigma_{\varepsilon v}}{(\beta - \alpha)^2}} = \beta + (\beta - \alpha)\cdot\frac{\sigma_{\varepsilon v} - \sigma_\varepsilon^2}{\sigma_v^2 + \sigma_\varepsilon^2 - 2\sigma_{\varepsilon v}}$$

如果 $\sigma_{\varepsilon v} = 0$，其经济涵义为影响需求和供给的因素是相互独立的，那么，

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



$$\text{plim}\hat{\beta}_{LS} = \beta + (\beta - \alpha) \cdot \frac{-\sigma_\varepsilon^2}{\sigma_v^2 + \sigma_\varepsilon^2}$$

$$\text{偏差} = (\beta - \alpha) \cdot \frac{-\sigma_\varepsilon^2}{\sigma_v^2 + \sigma_\varepsilon^2} > 0$$

由于$\beta < 0$且$\alpha > 0$，因此，$\beta - \alpha > 0$，使用普通最小二乘法得到的$\hat{\beta}_{LS}$要比实际的β大，即$\hat{\beta}_{LS} > \beta$。由于α与β相对称，同理可得：

$$\text{plim}\hat{\alpha}_{LS} = \alpha + (\beta - \alpha) \cdot \frac{-\sigma_v^2}{\sigma_v^2 + \sigma_\varepsilon^2}$$

$$\text{偏差} = (\beta - \alpha) \cdot \frac{-\sigma_v^2}{\sigma_v^2 + \sigma_\varepsilon^2} < 0$$

因此，使用普通最小二乘法得到的$\hat{\alpha}_{LS}$要比实际的α小，即：$\hat{\alpha}_{LS} < \alpha$。

在实际中，我们用价格p对销售量q进行回归时，不知道我们回归所得到的方程是需求方程还是供给方程。

举个例子来说，当回归系数为0.3时，我们应注意：

$\beta = 0.3 +$ 一个负数（即β肯定小于0.3）

$\alpha = 0.3 +$ 一个正数（即α肯定大于0.3）

当然，$\dfrac{\sum p_t q_t}{\sum p_t^2}$也有问题：当$\sigma_\varepsilon^2 \to 0$时，说明$\varepsilon$变化极小，而需求方程很稳定，变化不大。$\text{plim}\left(\dfrac{\sum p_t q_t}{\sum p_t^2}\right) = \beta$，如图12.6所示，而供给曲线变化明显，回归时我们得到的方程可能就是需求曲线方程。

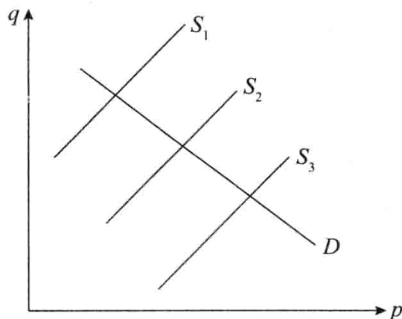


图 12.6

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

当 $\sigma_v \to 0$ 时，说明供给曲线十分稳定，$\text{plim}\left(\frac{\sum p_t q_t}{\sum p_t^2}\right) = \alpha$，如图 12.7 所示，而需求曲线变化明显，回归时所得到的方程就可能是供给曲线的方程。

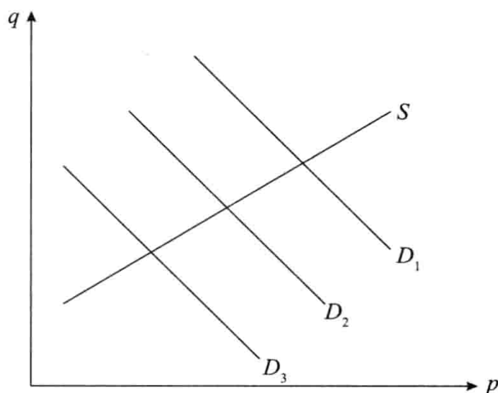


图 12.7

第八节　复习与提高

我们将第十章、第十一章、第十二章的复杂与提高并到一起来，是为了方便读者对比学习。在我们刚学过的三章中，主要是解决古典模型的六个前提遭到破坏时产生的问题。古典假设成立是完美的模型得以实现的先决条件，但是正因为抽象掉了现实中的一些矛盾使得古典模型的讨论显得偏离了现实。当现实与假设条件矛盾时，分别产生了异方差、自相关、联立方程组、多重共线性等问题，如何消除这些矛盾，使得在假设条件不成立时，我们仍能得到最好的、无偏的估计量，我们在这三章中看到，讨论的总思路仍是力图回到古典模型的过程中去：既然古典模型保证得到无偏最好估计量，那么我们就设法通过消去各种问题产生的偏差来产生一个新的古典模型。方法各有千秋，此处不再多述，表 12－6 中归纳了解决方法，请读者仔细阅读比较。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

表 12－6

古典假设条件	破坏	发现	贸然运用 OLS 结果	解决
1. $E(\varepsilon_i)=0$				
2. $\text{cov}(x_i,x_j)=0$	$\text{cov}(x_i,x_j)\neq 0$ 而 1、3、4、5、6 条件成立：多重共线性			
3. $Var(\varepsilon_i)=0$	$Var(\varepsilon_i)=\delta_i^2$ 而 1、2、4、5、6 条件成立：异方差	1. 图示法 2. ①Reset 检验：ε_i 对 $\hat{y}_i^1,\hat{y}_i^2,\hat{y}_i^3\cdots$ ②White 检验：ε_i 对 x_i 及其平方和交叉乘积 ③Glejser 检验 3. Goldfold－Quandt 检验	* 与 WLS 相比 $E(\hat{\beta}_{LS})=E(\hat{\beta}_{WLS})$ 无偏 $Var(\hat{\beta}_{LS})\geq Var(\hat{\beta}_{WLS})$ 无效 $E(\frac{S^2}{\sum x_i^2})\neq Var(\sigma^2)$	* 针对不同的 ε_i 形式，WLS 有不同解法 * 对 $y_i=\alpha+\beta x_i+\varepsilon_i$ $Var(\varepsilon_i)=\sigma^2(\alpha+\beta x)^2$ 两步 WLS * Deflation * 形式的选择 { *BM* 检验 { *PE* 检验
4. $\text{cov}(\varepsilon_i,\varepsilon_j)=0$	$\text{cov}(\varepsilon_i,\varepsilon_j)\neq 0$ 而 1、2、3、5、6 条件成立：自相关	1. DW 检验 2. Ident Resid 命令看是否为一阶自相关 3. 带前期变量 $y_t=\alpha y_{t-1}+\beta x_t+\varepsilon_t$ 时使用 Durbin－h 检验	* 计算机得到的估计值的方差 $Var(\hat{\beta}_{LS})$ 比真实值小，会盲目接受 T 检验与 F 检验 * 估计量无偏、无效 * 动态设定：①Sargan－Test②Walt－Test * 变量丢失：RESET、DW 检验均显著时加入变量而不用 AR（1） * 残差结构：Hendry＋Trivedi 检验	* 一阶差分方程 * GLS 法（ρ 已知时） ρ 未知时： { 循环查找 { 灰色查找 { 杜宾查找
5. $\text{cov}(\varepsilon_i,x_i)=0$	$\text{cov}(\varepsilon_i,x_i)\neq 0$ 而 1、2、3、4、6 条件成立：联立方程组	过度识别 不可识别	有偏估计 α_{LS} 对 α 低估 β_{LS} 对 β 高估 无法得出某一单个方程 单个方程中，β_{LS} 方差大于 TSLS 和 IV 的真值方差	* ILS 法 * IV 法（用于单个方程估计） * TSLS 法
6. 线性方程	非线性			

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

第十三章

基于时序动量回报率（Time – series Momentum Return）的量化交易策略

前面，我们系统地介绍了经济计量分析的基本方法，本章将通过具体例子来阐述基本的经济计量分析方法在量化投资中的使用。2011 年，国际著名的《金融经济学杂志》发表了芝加哥大学商学院 T. J. Moskowitz 教授和对冲基金 AQR 资本管理公司的 Yao Hua Ooi 以及纽约大学商学院 Lasse Heje Pederson 教授的《时序动量》一文。2012 年，此文获得对冲基金白盒（Whitebox）研究奖。

时序动量有别于截面动量（Cross – section Momentum）。在金融研究中，很多有关动量的研究大部分都是研究截面动量，也就是，比较在过去 3 到 12 个月里，相对于其它股票表现好的那些股票。与截面动量不同，时序动量不与其它股票进行截面比较，而是与其自身过去的回报率进行比较。

我们将采用时序动量收益率的概念，对上证综合指数 2005 年 9 月 2 日至 2013 年 9 月 2 日期间每周的历史数据进行分析，所使用的数据来源于雅虎财经网站。

我们需要使用 Excel 计算以下变量：

每周回报率（RT）=（本周收盘价 – 上周收盘价）/ 上周收盘价

前两周回报的标准差（Stdev）

时序动量汇报率（TSM – RT）= RT / Stdev

其结果如下：

表 13 – 1

2005 – 12 – 26	1161.06	RT	Stdev	TSM – RT
2006 – 1 – 2	1209.42	0.041652		
2006 – 1 – 9	1221.46	0.009955	0.022413	0.444175
2006 – 1 – 16	1255.31	0.027713	0.015886	1.744421

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

续前表

2005－12－26	1161.06	RT	Stdev	TSM－RT
2006－1－23	1258.05	0.002183	0.013086	0.166794
2006－1－30	1258.05	0	0.015409	0
2006－2－6	1282.66	0.019562	0.01072	1.82486
2006－2－13	1267.41	－0.01189	0.015881	－0.74866
2006－2－20	1296.87	0.023244	0.019309	1.203779
2006－2－27	1293.3	－0.00275	0.018229	－0.15101
2006－3－6	1245.65	－0.03684	0.030135	－1.22263
2006－3－13	1269.46	0.019115	0.028201	0.677802
2006－3－20	1294.7	0.019882	0.032531	0.611177
2006－3－27	1298.3	0.002781	0.00966	0.287851
2006－4－3	1342.96	0.034399	0.015827	2.173463
2006－4－10	1359.54	0.012346	0.016215	0.761389
2006－4－17	1416.79	0.04211	0.015447	2.726058
2006－4－24	1440.22	0.016537	0.016111	1.026457
……				

以上时序动量收益率（TSM－RT）在 EViews 中表示为（Y），使用 EViews 进行以下回归分析：

$$Y(t) = \beta * Y(t-1) + \xi$$

其运算结果如下：

Dependent Variable：Y				
Method：Least Squares				
Date：09/03/13　Time：21：39				
Sample（adjusted）：2 383				
Included observations：382 after adjusting endpoints				
Variable	Coefficient	Std. Error	t－Statistic	Prob.
Y（－1）	0.279014	0.049214	5.669413	0.0000
R－squared	0.074028	Mean dependentvar		0.103539
Adjusted R－squared	0.074028	S. D. dependentvar		1.621222
S. E. of regression	1.560060	Akaike info criterion		3.729940
Sum squaredresid	927.2731	Schwarz criterion		3.740268
Log likelihood	－711.4186	Durbin－Watson stat		2.119924

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

正如 Moskowitz 等人的结果一样，以上计算的 t – Statistic 结果说明，前一期时序动量收益率（Y（t – 1））对下一期时序动量收益率（Y（t））具有显著的影响，可以使用前一期的时序动量收益率（Y（t – 1））来设计交易信号。这里我们可以直接简单的 Moskowitz 时序量化策略，也就是说，现期时序动量收益率 Y（t）> 0 时，下期开始时买入；否则，下期空仓（假设只允许做多）。运用 Excel 来模拟此策略需要计算以下数列：

买入信号（Buysignal）：当前一期时序动量回报率（TSM – RT）> 0 时为 1；否则，为 0。

利润（Profit）：前一期利润 *（买入信号（Buysingal）* 回报率（RT）+1）。

基于 Excel 计算 Moskowitz 时序的量化策略结果见表 13 – 2。

表 13 – 2

	RT	Stdev	TSM – RT	buysignal	Profit
	0.041652				
	0.009955	0.022413	0.444175		
	0.027713	0.015886	1.744421		
	0.002183	0.013086	0.166794		
	0	0.015409	0		
	0.019562	0.01072	1.82486		
	–0.01189	0.015881	–0.74866		
	0.023244	0.019309	1.203779	0	1
	–0.00275	0.018229	–0.15101	1	0.997247
	–0.03684	0.030135	–1.22263	0	0.997247
	0.019115	0.028201	0.677802	0	0.997247
	0.019882	0.032531	0.611177	1	1.017075
	0.002781	0.00966	0.287851	1	1.019903
	0.034399	0.015827	2.173463	1	1.054986
……					

图 13.1 为简单的 Moskowitz 时序动能量化策略与上证指数的对比。从图中我们可以看出，以经过前期风险调整以后的时序动量回报率作为指标的利润曲线的波动幅度明显好于指数本身，尤其是最大回撤减少很多。尽管在 2008 年以后该策略的回报率一直好于指数，但是，在 2006 年到 2008 年初指数快速上涨的这段时间里，简单的时序动量策略的回报要远远低于指数本身。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

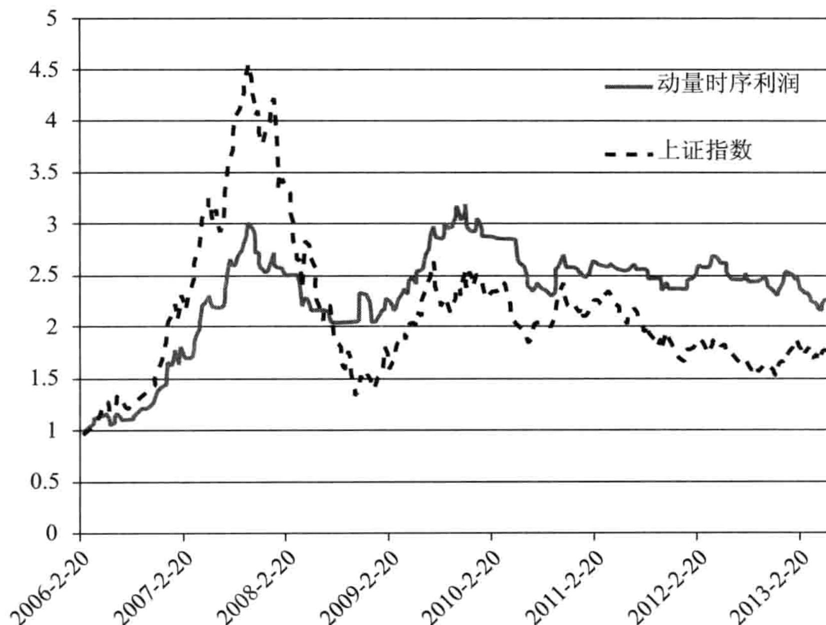


图 13.1

我们使用简单的移动平均方法，对简单的 Moskowitz 时序动量量化策略进行一些简单的改进。将前八期的时序动量回报率（TSM－RT）进行简单的平均，就在 EXCEL 中可以得到其前 8 期的均值数列 MA8。改进后的策略不再直接使用前一期的时序动量收益率来设计交易信号，而是使用 MA8 代替时序动量回报率（TSM－RT）作为信号，进行交易。也就是说，改进后的量化策略为：当 MA8 > 0 时，下期开始时买入；否则，下期空仓（假设只允许做多）。

运用 Excel 来模拟此改进后的策略，需要计算以下数列：

买入信号（Buysignal）：当前一期时序动量回报率 MA8 > 0 时，为 1；否则，为 0。

利润（Profit）：前一期利润 *（买入信号（Buysingal）* 回报率（RT）+1）。

表 13－3

RT	Stdev	TSM－RT	MA 8	buysignal	Profit
0.041652					
0.009955	0.022413	0.444175			
0.027713	0.015886	1.744421			
0.002183	0.013086	0.166794			

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

续前表

RT	Stdev	TSM - RT	MA 8	buysignal	Profit
0	0.015409	0			
0.019562	0.01072	1.82486			
-0.01189	0.015881	-0.74866			
0.023244	0.019309	1.203779	0.662196	0	1
-0.00275	0.018229	-0.15101	0.560545	1	0.997247
-0.03684	0.030135	-1.22263	0.362414	1	0.960505
0.019115	0.028201	0.677802	0.388372	1	0.978864
0.019882	0.032531	0.611177	0.262456	1	0.998327
……					

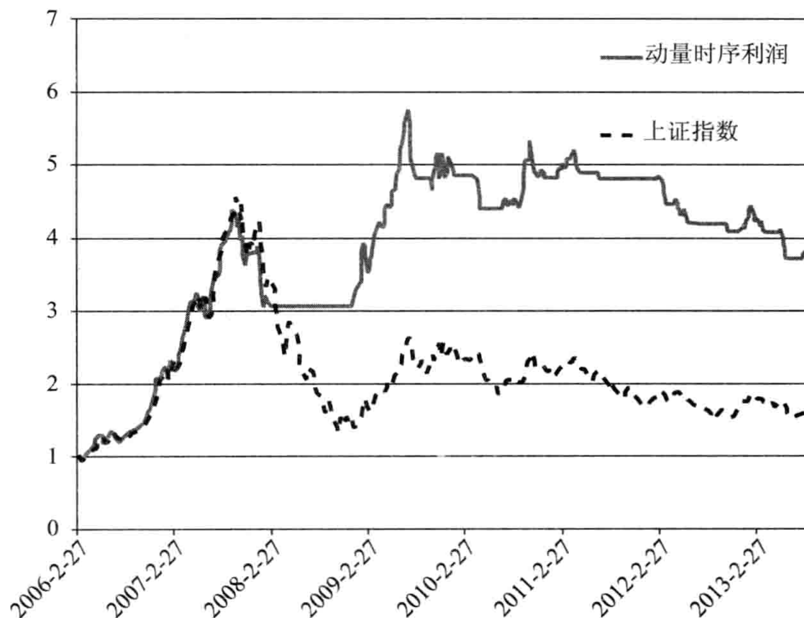


图 13.2

从图 13.2 可以看出，改进后的量化策略比简单 Moskowitz 量化策略有明显改进。2006 至 2008 年初的上升期，改进后的量化策略与指数几乎相差无几；在 2008 年以后，改进后的量化策略明显好于指数。尽管如此，改进后的量化策略在 2008 年以后不断小幅亏损的原因，主要是由于动量策略属于趋势跟踪类策略，所以对于没有趋势的盘整，只能不断的小幅亏损。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

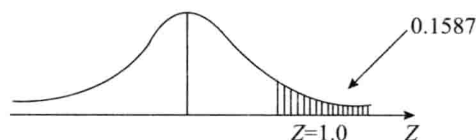
附 表

统计表

表 1　正态曲线下的面积

例　$z = \frac{x - u}{\sigma}$

$p(z > 1.0) = 0.1587$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2009	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1814	.1811	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

续前表

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0366	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

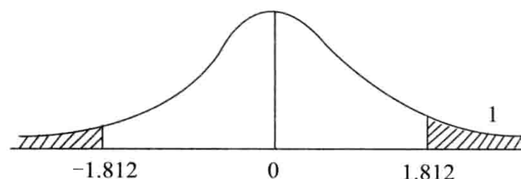
全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



表 2　t 分布的临界点

例：自由度 $v = 10, p(t > 1.812) = 0.05$

$p(t < -1.812) = 0.05$



a / v	.25	.20	.15	.10	.05	.025	.01	.005	.0005
1	1.000	1.376	1.963	3.076	6.314	12.706	31.821	63.657	636.619
2	.816	1.061	1.386	1.386	2.920	4.303	6.965	9.925	31.598
3	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841	12.941
4	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032	6.859
6	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405
8	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

续前表

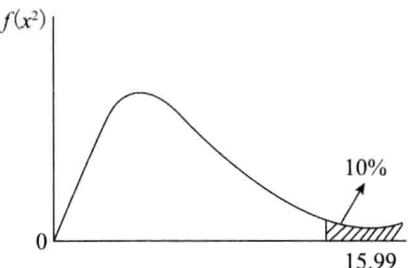
a / v	.25	.20	.15	.10	.05	.025	.01	.005	.0005
21	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	.858	1.060	1.319	1.714	2.096	2.500	2.807	3.767
24	.685	.857	1.059	1.318	1.711	2.064	2.492	2.397	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	.679	.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



表 3　x^2 分布的临界点

例：对于自由度 $v = 10, p(x^2 > 15.99) = 0.10$



v \ p	.955	.99	.975	.95	.90	.75	.50	.25	.10	.05	.025	.01	.005	p/v
1	0.0^4393	0.0^2157	0.0^2982	0.0^23	0.0158	0.102	0.455	1.323	2.71	3.84	5.02	6.63	7.88	1
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.386	2.77	4.61	5.99	7.38	9.21	10.60	2
3	0.0717	0.115	0.216	0.352	0.584	1.213	2.37	4.11	6.25	7.81	9.35	11.34	12.84	3
4	0.207	0.297	0.484	0.711	1.064	1.923	3.36	5.39	7.78	9.49	11.14	13.28	14.86	4
5	0.412	0.554	0.831	1.145	1.610	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	5
6	0.676	0.872	1.237	1.635	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	6
7	0.989	1.239	1.690	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.3	7
8	1.344	1.646	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.1	22.0	8
9	1.735	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.7	23.6	9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.5	23.2	25.2	10
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.9	24.7	26.8	11
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	11.85	18.55	21.0	23.3	26.2	28.3	12
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	10.81	22.4	24.7	27.7	29.8	13
14	4.07	4.66	5.63	6.57	7.79	1017	13.34	17.12	21.1	23.7	26.1	29.1	31.3	14
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.3	25.0	27.5	30.6	32.8	15

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

续前表

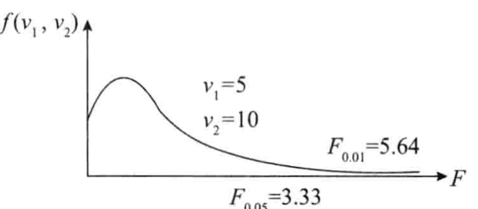
v \ p	.955	.99	.975	.95	.90	.75	.50	.25	.10	.05	.025	.01	.005	p/v
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.5	26.3	28.8	32.0	34.3	16
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.5	24.8	27.6	30.2	33.4	35.7	17
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.6	26.0	28.9	31.5	34.8	37.2	18
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.7	27.2	30.1	32.9	36.2	38.6	19
20	7.43	8.26	9.59	10.85	12.44	15.455	19.34	23.8	28.4	31.4	34.2	37.6	40.0	20
21	8.03	8.90	10.28	11.59	13.24	16.34	20.3	24.9	29.6	32.7	35.5	38.9	41.4	21
22	8.64	9.54	10.98	12.34	14.04	17.24	21.3	26.0	30.8	33.9	36.8	40.3	42.8	22
23	9.26	10.20	11.69	13.09	14.85	18.14	22.3	27.1	32.0	35.2	38.1	41.6	44.2	23
24	9.89	10.86	12.40	13.85	15.66	19.04	23.3	28.2	33.2	36.4	39.4	43.0	45.6	24
25	10.52	11.52	13.12	14.61	16.47	19.94	24.3	29.3	34.4	37.7	40.6	44.3	46.9	25
26	11.16	12.20	13.84	15.38	17.29	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.8	26
27	11.81	12.88	14.57	16.15	18.11	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	27
28	12.46	13.56	15.31	16.93	18.94	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	28
29	13.12	14.26	16.05	17.71	19.77	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	29
30	13.79	14.95	16.79	18.49	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	30
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	40
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5	50
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0	60
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	104.2	70
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.3	116.3	80
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3	90
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	140.2	100
Z_a	-2.58	-2.33	-1.96	-1.64	-1.28	-0.67	0.000	0.674	1.282	1.645	1.960	2.33	2.58	Z_a

注：当 $v>100$ 时取 $x2=\frac{1}{2}(za+\sqrt{2y-1})^2$，$Z_q$ 是相应显著水平 α 的标准正态偏差，列于表的最下一行。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



表 4　F 分布



例　自由度 $v_1 = 5, v_2 = 10, p(F > 3.33) = 0.05$.

$p(F > 5.64) = 0.01$

注：下面的黑体字是 1% 的显著水平，上面的为 5% 的显著水平。

分母自由度 v_2 \ 分子自由度 v_1	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6082	6106
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41
	98.49	99.00	99.17	99.25	99.30	99.33	99.34	99.36	99.38	99.40	99.41	99.42
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74
	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91
	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54	14.45	14.37
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68
	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.27	20.15	10.05	9.96	9.89
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.13	4.00
	13.74	10.92	9.78	9.15	8.75	8.47	8.25	8.10	7.98	7.87	7.79	7.72
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57
	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62	6.54	6.47
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28
	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82	5.74	5.67
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07
	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91
	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.73	4.71

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

续前表

分母自由度 v_2 \ 分子自由度 v_1	1	2	3	4	5	6	7	8	9	10	11	12
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79
	9.65	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69
	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60
	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53
	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48
	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42
	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28
	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25
	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23
	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20
	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18
	7.82	55.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



续前表

v_2 \ v_1	分子自由度 1	2	3	4	5	6	7	8	9	10	11	12
分母自由度 25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16
	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13
	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.14	3.06	2.98	2.93
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12
	7.64	5.45	4.57	4.07	3.76	3.53	3.36	3.23	3.11	3.03	2.95	2.90
29	4.18	3.33	3.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10
	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.08	3.00	2.92	2.87
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09
	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.90	2.84
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07
	7.50	5.34	4.46	3.97	3.66	3.42	3.25	3.12	3.01	2.94	2.86	2.80
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.50
	7.44	5.29	4.42	3.93	6.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03
	7.39	5.25	4.38	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99
	7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98
	7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97
	7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



续前表

分母自由度 v_2 \ v_1	分子自由度 1	2	3	4	5	6	7	8	9	10	11	12
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96
	7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95
	7.17	5.06	4.20	3.72	3.41	3.18	3.02	2.88	2.78	2.70	2.62	2.56
55	4.02	3.17	2.78	2.54	2.38/	2.27	2.18	2.11	2.05	2.00	1.97	1.93
	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90
	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.79	2.70	2.61	2.54	2.47
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89
	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88
	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.41
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85
	6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83
	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.65	2.56	2.47	2.40	2.33
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82
	6.81	4.75	3.91	3.44	3.14	2.92	2.76	2.62	2.53	2.44	2.37	2.30
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80
	2.39	4.71	3.88	3.41	3.11	2.90	2.73	2.60	2.50	2.41	2.34	2.28
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78
	6.70	4.66	3.83	3.36	3.06	2.85	2.69	2.55	2.46	2.37	2.29	2.23
1000	3.85	3.00	1.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76
	6.66	4.62	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.26	2.20
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75
	6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.24	2.18

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



续前表

v_2 \ v_1		分子自由度											
		14	16	20	24	30	40	50	75	100	200	500	∞
分母自由度	1	245	246	284	249	250	251	252	253	253	254	254	254
		6142	6169	6208	6234	6258	6286	6302	6323	6334	6352	6361	6366
	2	19. 42	19. 43	19. 44	19. 45	19. 46	19. 47	19. 47	19. 48	19. 49	19. 49	19. 50	19. 50
		99. 43	99. 44	99. 45	99. 46	99. 47	99. 48	99. 48	99. 49	99. 49	99. 49	99. 50	99. 50
	3	8. 71	8. 69	8. 66	8. 64	8. 62	8. 60	8. 58	8. 57	8. 56	8. 54	8. 53	8. 53
		26. 92	26. 83	26. 69	26. 60	26. 50	26. 41	26. 35	26. 27	26. 23	26. 18	26. 14	26. 12
	4	5. 87	5. 84	5. 80	5. 77	5. 74	5. 71	5. 70	5. 68	5. 66	5. 65	5. 64	5. 63
		14. 24	14. 15	14. 02	13. 93	13. 83	13. 74	13. 69	13. 61	13. 57	13. 52	13. 48	13. 46
	5	4. 64	4. 60	4. 56	4. 53	4. 50	4. 46	4. 44	4. 42	4. 40	4. 38	4. 37	4. 36
		9. 77	9. 68	9. 55	9. 47	9. 38	9. 29	9. 24	9. 17	9. 13	9. 07	9. 04	9. 02
	6	3. 96	3. 92	3. 87	3. 84	3. 81	3. 77	3. 75	3. 72	3. 71	3. 69	3. 68	3. 67
		7. 60	7. 52	7. 33	7. 31	7. 23	7. 14	7. 09	7. 02	6. 99	6. 94	6. 90	6. 88
	7	3. 52	3. 49	3. 44	3. 41	3. 38	3. 34	3. 32	3. 29	3. 28	3. 25	3. 24	3. 23
		6. 35	6. 27	6. 15	6. 07	5. 98	5. 90	5. 85	5. 78	5. 75	5. 70	5. 67	5. 65
	8	3. 23	3. 20	3. 15	3. 12	3. 08	3. 05	3. 03	3. 00	2. 98	2. 96	2. 94	2. 93
		5. 56	5. 48	5. 36	5. 28	5. 20	5. 11	5. 06	5. 00	4. 96	4. 91	4. 88	4. 86
	9	3. 02	2. 98	2. 93	2. 90	2. 86	2. 82	2. 80	2. 77	2. 76	2. 73	2. 72	2. 71
		5. 00	4. 92	4. 80	4. 73	4. 64	4. 56	4. 51	4. 45	4. 41	4. 36	4. 33	4. 31
	10	2. 86	2. 82	2. 77	2. 74	2. 70	2. 67	2. 64	2. 61	2. 59	2. 56	2. 55	2. 54
		4. 60	4. 52	4. 41	4. 33	4. 25	4. 17	4. 12	4. 05	4. 01	3. 96	3. 93	3. 91
	11	2. 74	2. 70	2. 65	2. 61	2. 57	2. 53	2. 50	2. 47	2. 45	2. 42	2. 41	2. 40
		4. 29	4. 21	4. 10	4. 02	3. 94	3. 86	3. 80	3. 74	3. 70	3. 66	3. 62	3. 60
	12	2. 64	2. 60	2. 54	2. 50	2. 46	2. 42	2. 40	2. 36	2. 35	2. 32	2. 31	2. 30
		4. 05	3. 98	3. 86	3. 78	3. 70	3. 61	3. 56	3. 49	3. 46	3. 41	3. 38	3. 36
	13	2. 55	2. 51	2. 46	2. 42	2. 38	2. 34	2. 32	2. 28	2. 26	2. 24	2. 22	2. 21
		3. 85	3. 78	3. 67	3. 59	3. 15	3. 42	3. 37	3. 30	3. 27	3. 21	3. 18	3. 16

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

续前表

v_2 \ v_1	分子自由度											
分母自由度	14	16	20	24	30	40	50	75	100	200	500	∞
14	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13
	3.70	3.62	3.51	3.43	3.34	3.26	3.21	3.14	3.11	3.06	3.02	3.00
15	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07
	3.56	3.48	3.36	3.29	3.20	3.12	3.07	3.00	2.97	2.92	2.89	2.87
16	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01
	3.45	3.37	3.25	3.18	3.10	3.01	2.96	2.89	2.86	2.80	2.77	2.75
17	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96
	3.35	3.27	3.16	3.08	3.00	2.92	2.86	2.79	2.76	2.70	2.67	2.65
18	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92
	3.27	3.19	3.07	3.00	2.91	2.83	2.78	2.71	2.68	2.62	2.59	2.57
19	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88
	3.19	3.12	3.00	2.92	2.84	2.76	2.70	2.63	2.60	2.54	2.51	2.49
20	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84
	3.13	3.05	2.94	2.86	2.77	2.69	2.63	2.56	2.53	2.47	2.44	2.42
21	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81
	3.07	2.99	2.88	2.80	2.72	2.63	2.58	2.51	2.47	2.42	2.38	2.36
22	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78
	3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.37	2.33	2.31
23	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76
	2.97	2.89	2.78	2.79	2.62	2.53	2.48	2.41	2.37	2.32	2.28	2.26
24	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73
	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36	2.33	2.27	2.23	2.21
25	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71
	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.32	2.29	2.23	2.19	2.17
26	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69
	2.86	2.77	2.66	2.58	2.50	2.41	2.36	2.28	2.25	2.19	2.15	2.13

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net



续前表

v_2 \ v_1		分子自由度											
		14	16	20	24	30	40	50	75	100	200	500	∞
分母自由度	27	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67
		2.83	2.74	2.63	2.55	2.47	2.38	2.33	2.25	2.21	2.16	2.12	2.10
	28	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65
		2.80	2.71	2.60	2.52	2.44	2.35	2.30	2.22	2.18	2.13	2.09	2.06
	29	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64
		2.77	2.68	2.57	2.49	2.41	2.32	2.27	2.19	2.15	2.10	2.06	2.03
	30	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62
		2.74	2.66	2.55	2.47	2.38	2.29	2.24	2.16	2.13	2.07	2.03	2.01
	32	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59
		2.70	2.62	2.51	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96
	34	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57
		2.66	2.58	2.47	2.38	2.30	2.21	2.15	2.08	2.04	1.98	1.94	1.91
	36	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55
		2.62	3.54	2.43	2.35	2.26	2.17	2.12	2.04	2.00	1.94	1.90	1.87
	38	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53
		2.59	2.51	2.40	2.32	2.22	2.14	2.08	2.00	1.97	1.90	1.86	1.84
	40	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
		2.56	2.49	2.37	2.29	2.20	2.11	2.05	1.97	1.94	1.88	1.84	1.81
	42	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49
		2.54	2.46	2.35	2.26	2.17	2.08	2.02	1.94	1.91	1.85	1.80	1.78
	44	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48
		2.52	2.44	2.32	2.24	2.15	2.06	2.00	1.92	1.88	1.82	1.78	1.75
	46	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46
		2.50	2.42	2.30	2.22	2.13	2.04	1.98	1.90	1.86	1.80	1.76	1.72
	48	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45
		2.48	2.40	2.28	2.20	2.11	2.02	1.96	1.88	1.84	1.78	1.73	1.70

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

续前表

分母自由度 v_2 \ 分子自由度 v_1	14	16	20	24	30	40	50	75	100	200	500	∞
50	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44
	2.46	2.39	2.26	2.18	2.10	2.00	1.94	1.86	1.82	1.76	1.71	1.68
55	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41
	2.43	2.35	2.23	2.15	2.06	1.96	1.90	1.82	1.78	1.71	1.66	1.64
60	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39
	2.40	2.32	2.20	2.12	2.03	1.93	1.87	1.79	1.74	1.68	1.63	1.60
65	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37
	2.37	2.30	2.18	2.09	2.00	1.90	1.84	1.76	1.71	1.64	1.60	1.56
70	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35
	2.35	2.28	2.15	2.07	1.98	1.88	1.82	1.74	1.69	1.62	1.56	1.53
80	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32
	2.32	2.24	2.11	2.03	1.94	1.84	1.78	1.70	1.65	1.57	1.52	1.49
100	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28
	2.26	2.19	2.06	1.98	1.89	1.79	1.73	1.64	1.59	1.51	1.46	1.43
125	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25
	2.23	2.15	2.03	1.94	1.85	1.75	1.68	1.59	1.54	1.46	1.40	1.37
150	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22
	2.20	2.12	2.00	1.91	1.83	1.72	1.66	1.56	1.51	1.43	1.37	1.33
200	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19
	2.17	2.09	1.97	1.88	1.79	1.69	1.62	1.53	1.48	1.39	1.33	1.28
400	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13
	2.12	2.04	1.92	1.84	1.74	1.64	1.57	1.47	1.42	1.32	1.24	1.19
1000	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08
	2.09	2.01	1.89	1.81	1.71	1.61	1.54	1.44	1.38	1.28	1.19	1.11
∞	1.67	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00
	2.07	1.99	1.87	1.79	1.69	1.59	1.52	1.41	1.36	1.25	1.15	1.00

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

表 5 杜宾—瓦特森检验上下界

5%的上下界

n	$k=2$		$k=3$		$k=4$		$k=5$		$k=6$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.81
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

续前表　　　　1% 的上下界

n	$k=2$		$k=3$		$k=4$		$k=5$		$k=6$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.27	0.74	1.41	0.65	1.58	0.56	1.74
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.53	1.35	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

注：n 是观察值的数目。
　　k 是解释变量的数目，包括常数项。

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

参考书目

[1] Fredman D. , Pisani R. and Purves R. Statistics (4th Edition) . W. W. Norton & Company, 2007.

[2] H. W. Greene. Econometric Analysis (6th edition) . Prentice Hall, 2007.

[3] A. S. Goldberger. Econometric Theory. New York: Wiley, 1964.

[4] Intriligator M. D. Econometric Models, Techniques and Application (2nd Edition) . Prentice – Hall. 1995.

[5] Johnston J. Econometric Methods (4th Edition) . McGraw-Hill Higher Education, 1997.

[6] Peter Kennedy. A Guide to Econometrics (6th Edition) . Wiley-Blackwell, 2008.

[7] G. S. Maddala. Introduction to Econometrics (3rd Edition) . Wiley, 2001.

[8] Pindyck R. S. and Rubinfeld D. L. Econometric Models and Economic Forecasts (4th Edition) . McGraw-Hill Higher Education, 2000.

[9] Wooldridge J. Introductory Econometrics: A Modern Approach (3rd Edition) . South-Western College Pub, 2005.

[10] 刘振亚编著．计量经济学教程．北京：中国人民大学出版社，1997.

[11] 古扎拉蒂．计量经济学基础（第四版）．北京：中国人民大学出版社，2005.

[12] 贾俊平，何晓群，金勇进编著．统计学．北京：中国人民大学出版社，2000.

[13] 劳伦斯·克莱因，理查德·扬．经济计量预测与预测模型入门．北京：中国社会科学出版社，1982.

[14] 龙永红主编．概率论与数理统计．北京：中国人民大学出版社，2004.

[15] 邹至庄．用控制方法进行计量经济分析．北京：中国友谊出版社，1987.

[16] 张晓峒．计量经济学基础．天津：南开大学出版社，2007.

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服微信/QQ：958218088 期货开户和书籍下载请进:http://www.7help.net

刘振亚教授主编的量化投资方法丛书，一定会对中国基金业未来发展注入新的活力！

——中国证券投资基金业协会会长 孙杰

量化投资是现代资产管理的重要手段。随着中国证券和期货市场的逐步成熟，量化投资的理念和方法必将成为市场的主流之一。

——摩根大通期货有限公司董事长 周小雄

刘振亚教授从事量化投资研究和实践多年，终于将研究成果付梓，由衷地感到高兴!

——中国人民大学金融与证券研究所所长 吴晓求

上架建议 金融投资

ISBN 978-7-5136-283



定价：36.00元